Implicit Social Cognition:

Evidence from Subliminal Priming Experiments and Implicit Association Tests (IATs)

Thesis (cumulative thesis)
presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the degree of Doctor of Philosophy

by Livio Raccuia

Accepted in the spring semester 2016
on the recommendation of the doctoral committee:
Prof. Dr. Marco Steenbergen (main supervisor)
Prof. Dr. Dominik Hangartner

Zurich, 2016

Abstract

Implicit social cognition is the study of implicit (i.e., automatic or even unconscious) stereotypes and attitudes. Although it is a relatively young field of research, implicit social cognition has already left its mark on the discipline of social psychology. In fact, there is now a near-consensus among psychologists that implicit stereotypes and attitudes matter for the well-being of many stigmatized groups. In addition, a more recent discussion in the field has dealt with the utility of implicit measures for the prediction of voting behavior among undecided voters. Other than explicit (i.e., self-reported) measures, implicit measures were seen as fit instruments to read the minds of those who are unwilling or unable to report their voting intention.

In this dissertation, I aim to set the stage for research on implicit social and political cognition in Switzerland. Unfortunately, research on implicit social cognition has not yet set foot on Switzerland, at least not among political scientists. This is very unfortunate, for the country's direct democratic system provides many opportunities to run studies on implicit social (political) cognition. For example, the recent, and still ongoing, discussion about the efficacy of political fear campaigns suggests a closer examination of the authoritarianism under threat hypothesis. Research on right-wing authoritarianism has shown that authoritarian predispositions can be triggered upon exposure to societal threat. In addition, consumer psychologists have demonstrated that subliminal (i.e., unconscious) ads may influence the behavior of individuals whose motivations are addressed by the subliminal prime. Building on these insights, I examine whether the subliminal exposure to right-populist poster ads affects the high authoritarian segment of the population. In both a laboratory and web experiment, I find that high authoritarian individuals become more likely to sign a petition against foreign students or support a ban on mosques and minarets as a function of the subliminal poster ads. It thus seems that conservative or right-populist parties have a secret tool to persuade their followers.

In the second chapter, I address the debate on the predictive value of implicit attitudes for the prediction of voting behavior. More concretely, I report results from three studies that include both implicit and explicit measures. All three studies were conducted ahead of Swiss federal votes. I find that implicit and explicit attitudes are equally good predictors among undecided voters while explicit attitudes are better predictors among decided voters. Furthermore, implicit attitudes predict voting behavior descriptively, but not significantly, better for undecided voters. Thus, my results suggest that, as was argued in previous research, the utility of implicit attitudes may be higher in the context of voting on political issues rather than elections, but still not as high as initially believed or hoped-for.

Finally, in the third chapter, we replicate previous research on the detrimental effects of subtle gender bias on female students' chances to pursue academic science careers. Using

the recently proposed parallel encouragement design (PED) and Amazon Mechanical Turk (MTurk), we do not find negative effects of female gender on perceived hireability for a lab manager position or salary offer. Instead, we find that subjects would offer more starting salary to female students largely as a function of their greater likeability. Moreover, we find only weak implicit gender stereotypes among study participants, and implicit stereotypes do not moderate the indirect effect between gender, perceived likeability and salary offer.

Taken together, my results suggest that implicit social cognition does indeed matter, but only under very specific conditions and sometimes not as much as purported in the literature.

Contents

Abstract **i**

List of Figures **vii**

List of Tables **viii**

Acknowledgments **xii**

Preface **xiv**

A Synopsis **1**

1 Implicit Social (Political) Cognition **2**
- 1.1 Introduction . . . 2
- 1.2 What is Implicit Social (Political) Cognition and Where Does It Come from? . . . 3
- 1.3 Implicit Attitude Measures . . . 5
- 1.4 Does Implicit Social (Political) Cognition Really Matter? . . . 9
- 1.5 Motivation and Background of this Dissertation . . . 11
- 1.6 Results and Discussion . . . 12

References **15**

B Articles **21**

1 Subliminal Fear Ads and Right-wing Authoritarianism: Experimental Evidence from Switzerland and the US **22**
- 1.1 Introduction . . . 23
- 1.2 Literature Review and Theory . . . 24
  - 1.2.1 Subliminal advertising . . . 24
  - 1.2.2 Authoritarianism (under threat) . . . 26
- 1.3 Theoretical Model . . . 27
- 1.4 Study 1 . . . 27
  - 1.4.1 Participants and design . . . 27
  - 1.4.2 Procedure . . . 28
  - 1.4.3 Independent variables . . . 29

1.4.4 Dependent variables . . . 30
1.4.5 Results and discussion . . . 30
1.5 Study 2 . . . 33
1.5.1 Participants and design . . . 33
1.5.2 Procedure . . . 33
1.5.3 Independent variables . . . 35
1.5.4 Dependent variables . . . 35
1.5.5 Results and discussion . . . 35
1.6 General Discussion . . . 37

References **39**

Appendix **43**
1.A Stimuli List for Subliminal Priming Task . . . 43

2 Single-Target Implicit Association Tests (ST-IAT) Predict Voting Behavior of Decided and Undecided Voters in Swiss Referendums **44**
2.1 Introduction . . . 45
2.2 Implicit Attitude Measures and the Prediction of Political Behavior . . . 46
2.3 Study 1 . . . 48
2.3.1 Ethics statement . . . 48
2.3.2 Participants . . . 49
2.3.3 Procedure . . . 49
2.3.4 Measures . . . 49
2.3.5 Results and discussion . . . 51
2.4 Study 2 . . . 54
2.4.1 Ethics statement . . . 54
2.4.2 Participants . . . 54
2.4.3 Procedure . . . 54
2.4.4 Measures . . . 54
2.4.5 Results and discussion . . . 56
2.5 Study 3 . . . 58
2.5.1 Ethics statement . . . 59
2.5.2 Participants . . . 59
2.5.3 Procedure . . . 59
2.5.4 Measures . . . 60
2.5.5 Results and discussion . . . 60
2.6 General Discussion . . . 66

References **69**

Appendix **72**
2.A Stimuli List for ST-IATs on Minimum Wage Initiative & Gripen Referendum (Study 1) . . . 72
2.B First condition of computer-administered paper-format ST-IAT (Study 2) . . . 73
2.C Stimuli List for ST-IAT on Public Health Insurance Initiative (Study 2) . . . 74
2.D Stimuli List for ST-IAT on Ecopop Initiative (Study 3) . . . 75
2.E Results of linear regression of the explicit measure on the implicit measure, cognitive elaboration, and their interaction . . . 76

3 The Causal Mediation Effect of Likeability on Female Students' Career Prospects in Science **77**
3.1 Introduction . . . 78
3.2 Experimental Design . . . 82
3.3 Results . . . 84
3.3.1 Student gender differences . . . 84
3.3.2 Mediation analyses . . . 85
3.4 Discussion . . . 89

References **91**

Appendix **95**
3.A Introduction . . . 95
3.B Participants and Recruitment . . . 95
3.C Student Laboratory Manager Application Materials . . . 96
3.D Encouragements . . . 98
3.E Measures . . . 99
3.F Analytic Strategy . . . 103
3.G Sensitivity Analyses . . . 106

List of Figures

A Synopsis **2**

1.1 Illustration of the scientific contextualization of the dissertation 3

B Articles **22**

1.1 Posters Used in Forced-Choice Test of Study 1 29
1.2 Results for Petition for Foreign Student Quota, Study 1 32
1.3 Prime and Mask, Study 2 34
1.4 Posters Used in Forced-Choice Test of Study 2 34
1.5 Results for Ban on Mosques and Minarets, Study 2 37
1.A.1 Pleasant pictures used in subliminal priming task (Study 2) 43
1.A.2 Unpleasant pictures used in subliminal priming task (Study 2) 43

2.1 Probability of voting for Ecopop as a function of implicit attitudes, cognitive elaboration, and their interaction (left), and as a function of explicit attitudes, cognitive elaboration, and their interaction (right) 65

3.1 The causal mechanism between female gender identity, perceived competence, and perceived student hireability and salary conferral 79
3.2 Structure of the parallel encouragement design (PED) 81
3.3 Randomized encouragement as an instrument for the mediator 82
3.4 Graphical summary of the average mediation effect (ACME), average direct effect (ADE), and the total effect of student gender on perceived hireability (left panel) and salary conferral (right panel) 88
3.C.1 Cover story text 97
3.C.2 Lab manager application materials for one of the female student conditions . . 98
3.D.1 News articles in the [low/high] encouragement conditions of the second experiment (D=1) 99
3.E.1 Screenshot of the IAT 101
3.G.1 Sensitivity analyses for the causal mechanism between student gender, likeability and perceived hireability (left) and salary conferral (right) 108

List of Tables

A Synopsis **2**

1.1 Sequence of Trial Blocks in the Classic IAT 6
1.2 Overview of Implicit Measures 8

B Articles **22**

1.1 Descriptive statistics and correlations between right-wing authoritarianism and the dependent variables (Study 1) 31
1.2 Dependent variables as a function of priming condition and authoritarianism (Study 1) 31
1.3 Descriptive statistics and correlations between right-wing authoritarianism and the dependent variable (Study 2) 36
1.4 Support for a ban on mosques and minarets as a function of priming condition and right-wing authoritarianism (Study 2) 36

2.1 Sequence of Trial Blocks in ST-IATs in Study 1 50
2.2 Results of logistic regression for prediction of voting intention (Study 1) 52
2.3 Results of logistic regression for prediction of voting behavior (Study 1) 53
2.4 Results of logistic regression for prediction of voting intention (Study 2) 57
2.5 Results of logistic regression for prediction of voting behavior (Study 2) 58
2.6 Results of logistic regression for prediction of voting intention (Study 3) 61
2.7 Results of logistic regression for prediction of voting behavior, separately for decided and undecided voters (Study 3) 63
2.8 Results of logistic regression analyses predicting voting behavior from explicit (EA) and implicit attitudes (IA) and decidedness (Study 3) 64
2.9 Results of multiple binary logistic regression analyses for the prediction of voting behavior including cognitive elaboration as moderator (Study 3) 65
2.A.1 Stimuli List for ST-IATs on Minimum Wage Initiative & Gripen Referendum (Study 1) 72
2.B.1 First condition of computer-administered paper-format ST-IAT (Study 2) . . 73
2.C.1 Stimuli List for ST-IAT on Public Health Insurance Initiative (Study 2) 74
2.D.1 Stimuli List for ST-IAT on Ecopop Initiative (Study 3) 75

2.E.1 Results of linear regression of the explicit measure on the implicit measure, cognitive elaboration, and their interaction 76

3.1 Means for student competence, hireability, and salary conferral by student gender condition and participant gender 84
3.2 Estimated causal mediation effects from 2SLS regressions 86
3.3 Estimated causal mediation effects of female student gender and perceived competence/likeability 87
3.B.1 Descriptive statistics for subjects in different experimental conditions 96
3.C.1 Treatment Stimuli Sets 96
3.E.1 Sequence of trial blocks in Gender-Science IAT 100
3.E.2 Stimuli used in Gender-Science IAT 101
3.F.1 First Stage Regression Estimates 104
3.F.2 2SLS Estimates of the Effect of Female Student Gender on Perceived Hireability and Salary Conferral 105

To my parents.

Acknowledgments

FINISHING a dissertation is a big achievement. Not only because one needs to invest a lot of time in the reading and writing of articles, the design of surveys and experiments, the collection and analysis of data, but also because of the tremendous amount of time and energy necessary to get the research going, recover from setbacks, and give it another shot. As I am writing this, I realize just how many people have been involved in this thesis and contributed to its realization. I sincerely hope not to forget anyone, please forgive me in case I still do.

First and foremost, I would like to thank my two supervisors Prof. Dr. Marco Steenbergen and Prof. Dr. Dominik Hangartner for their valuable inputs and thoughtful suggestions. Throughout my doctorate, Marco granted me the outmost autonomy and intellectual freedom. He always offered me good advice and funding for my studies. Dominik has been a good friend and mentor ever since I took his course on causal inference in fall 2009. He is a very inspiring and caring person and I am glad he convinced me of doing a Ph.D. Thank you Marco and Dominik for giving me the opportunity to work on a topic of personal interest and for believing in my abilities to complete a Ph.D.

Parts of my research were funded by the Foundation for Research in Science and the Humanities at the University of Zurich and the Foundation for Migration, Population, and Environment. Without their generous support, the present dissertation would not have been possible. My dissertation has also benefited from excellent research assistance by Christian Müller, Stuart-Unwin Smith, and Stefan Wehrli. In particular, Christian has been a very reliable and constant source of help. I am also grateful for discussions with Prof. Dr. Werner Wirth in 2012 and Prof. Dr. Johannes Ullrich in 2015. Their critical comments and suggestions clearly improved my dissertation. I would also like to thank participants of the 2014 Midwest Political Science Association (MPSA) conference in Chicago, participants of the 2015 European Political Science Association (EPSA) conference in Vienna, participants of the 2015 International Society of Political Psychology (ISPP) conference in San Diego, and participants of the 2015 IBH workshop in Constance (Germany) for their helpful comments and suggestions.

During my Ph.D., I received support from many dear friends at the Chair of Methods: Marlène Gerber, Selina Kurer, Fabian Morgenthaler, Seraina Pedrini, Giuseppe Pietrantuono, Livia Schubiger, Stefan Schütz, Tomasz "Tomek" Siczek, Judith Spirig, Kushtrim Veseli, and Edward Weber have shared many unforgettable moments with me. Many of them showed up and cheered for me at the Natural Bodybuilding Swiss Championships in 2014 and 2015. A big shout-out to all of you for the ongoing support and friendship! On this occassion, I shall not forget my close friend and "coach" Clint Claessen whom I got to know through teaching courses of the MA CIS program. Clint was an exchange student and he was more impressed

by my gains than my teaching skills. Anyway, I am proud to call you my friend.

Finally, I would like to thank my parents Franziska and Marino for their endless patience and for teaching me important virtues that have undoubtedly contributed to the successful completion of this dissertation. Of course, my girlfriend Amina also deserves special mention for her constant support and love. You have the warmest heart and I respect and love you for so many things. Thank you!

Preface

This dissertation comes in two main parts: In the first part (A), I discuss the origins and relevance of research on implicit social and political cognition before addressing the motivation and background of my own research. The chapter is rounded up by a brief summary and discussion of my results. In the second part (B), I present the three papers that constitute my cumulative dissertation. The three papers are:

1. Raccuia, Livio. 2016. "Subliminal Fear Ads and Right-Wing Authoritarianism: Experimental Evidence from Switzerland and the US." Manuscript.

2. Raccuia, Livio. 2016. "Single-Target Implicit Association Tests (ST-IAT) Predict Voting Behavior of Decided and Undecided Voters in Swiss Referendums." *Public Library of Sciences ONE (PLoS ONE).* 2016; 11(10): e0163872. doi: 10.1371/journal.pone.0163872.

3. Raccuia, Livio and Marco Steenbergen. 2016. "The Causal Mediation Effect of Likeability on Female Students' Career Prospects in Science." Manuscript.

Part A

Synopsis

1

Implicit Social (Political) Cognition

1.1 Introduction

With the US presidential election around the corner, let me start by asking the following question: Have you ever wondered what exactly you like so much about a particular political candidate even though she is from your least favorite party? Despite your brain telling you not to vote for her, you somehow cannot help but like her. Implicit social cognition may have an answer to your question. In addition, it may be able to explain why most people get irritated by the use of female pronouns. Are not most political candidates men (apart from this year's election, of course)?

Implicit social cognition is the study of gut feelings or, in more scientific terms, implicit attitudes and stereotypes. As one of the fastest growing fields in psychology, implicit social cognition has given rise to a multitude of implicit measures and a confusing glossary. For this reason, I will briefly address the origins of the field and provide an intuitive definition of implicit social cognition in the next section (1.2). I will then (1.3) introduce some of the most frequently used implicit attitude measures, with a particular focus on the implicit association test (IAT). In the fourth section (1.4), I will provide a summary of the most important insights that research on implicit social cognition has produced. In the fifth section (1.5), I will discuss the particular motivation for writing this dissertation before discussing my results in the sixth section (1.6).

1.2 What is Implicit Social (Political) Cognition and Where Does It Come from?

Undoubtedly, the theoretical foundations of implicit social cognition could easily be traced back as far as Freud's psychoanalytic theory or Helmholtz's (1910, 1925) suggestion that many mental processes occur outside of conscious awareness. Indeed, Greenwald and Banaji's (1995) definition of implicit social cognition draws heavily from previous research on implicit memory and its focus on how past and unconscious experience can influence later performance. According to them, implicit social cognition is the study of cognitive processes that occur outside of conscious awareness or conscious control. More intuitively, implicit social cognition is about studying the effects of social cognitions such as attitudes or stereotypes that individuals may not be aware of. It follows that implicit political cognition is a narrower field of research, primarily concerned with implicit political attitudes and stereotypes. For example, the first two papers of this dissertation are concerned with the effects of subliminal and thus unconscious exposure to political poster ads (see chapter B1) and implicit attitudes toward political referendums (see chapter B2). The third paper deals with the effects of implicit gender stereotypes on women's chances of pursuing academic science careers. Hence, it represents a piece of research that is more typical for implicit social cognition (see Figure 1.1).

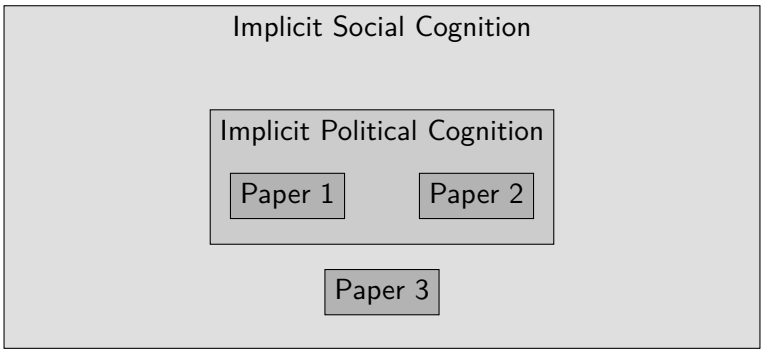

Figure 1.1: Illustration of the scientific contextualization of the dissertation

However, from a contemporary point of view, the above definition of implicit social (political) cognition is incomplete and even somewhat misleading. In fact, recent research has blurred the line on whether implicit attitudes truly operate below conscious awareness. For example, Gawronski et al. (2006) have reviewed previous research on implicit attitudes with

respect to three particular aspects of unconsciousness: (i) Awareness about the source or cause of an attitude (source awareness), (ii) awareness about the content of an attitude (content awareness), and (iii) awareness about the impact of an attitude (impact awareness). They conclude that in most studies, people were consciously aware of their implicit attitudes (content awareness) but not of their influence on other psychological processes (impact awareness). Furthermore, they note that lack of awareness about the cause or origin of an attitude (source awareness) could be observed for both implicit and explicit (i.e., self-reported) measures. In a similar vein, Hahn et al. (2014) have demonstrated that people can accurately predict their performance on implicit attitude measures and thus are aware of their implicit attitudes. It thus seems that the unique feature of implicit cognitions is not individuals' lack of content or source awareness but rather people's inability to anticipate the effects that implicit attitudes and stereotypes may have on their judgments and behavior.

What, then, are implicit attitudes and stereotypes? According to Greenwald and Banaji (1995, 8), implicit attitudes are ...

> "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects."

In other words, implicit attitudes are associations (memories) with a particular attitude object (e.g., political party) that an individual is not necessarily aware of and whose impact on subsequent behavior she cannot consciously control. Similarly, implicit stereotypes are ...

> "introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category" (Greenwald and Banaji 1995, 15).

Since the ascribed qualities can be both positive and negative, implicit stereotypes (much like stereotypes in general) need not necessarily result in the downgrading of social groups. However, if the traits that are associated with a particular group are predominantly negative, implicit stereotypes may eventually result in implicit bias and discrimination. For example, research on implicit racial bias (e.g., Devine 1989) has repeatedly demonstrated that African Americans are associated with both positive (e.g., athletic) and negative (e.g., lazy) qualities, but also that the latter clearly outweigh the former. As a consequence, implicit racial stereotypes can lead to discrimination in the job market (e.g., Ziegert and Hanges 2005) or significantly higher probabilities of being mistakenly shot by police officers (e.g., Correll et al. 2002).

These preliminary (see also section 1.4) examples show that implicit social cognitions can have profound effects on the lives of stigmatized social groups. It goes without saying, however, that any study of implicit social or political cognition requires the analyst to measure the implicit attitude or stereotype of interest somehow. To this end, a large number of implicit measures have been developed. I shall address these measures in the next section.

1.3 Implicit Attitude Measures

In the 1960s and 1970s, cognitive psychologists (e.g., Bruner 1957; Meyer and Schvanefeldt 1971, 1976; Neely 1977; Rosch 1975) proposed an associationist theory of memory. Mental constructs were believed to be connected with each other such that activation of one construct (e.g., "Obama") would automatically activate associated constructs (e.g., "president"). According to this research, automatic activation should primarily occur for strongly connected items, while loosely connected constructs were believed to require more effortful (i.e., controlled) processing. Building on these theoretical underpinnings, social psychologists (e.g., Dovidio et al. 1986; Gaertner and McLaughlin 1983) later demonstrated the automatic activation of attitudes using the sequential priming procedure. In this approach, the short presentation of a prime (e.g., Black face) is followed by a target word (e.g., "wonderful"). Subjects are asked to sort these words (or pictures) into specific categories (e.g., good and bad) as fast and accurately as possible. The fundamental idea behind the sequential priming technique is that correspondence between a subject's true attitude and the respective prime-target sequence (e.g., Black-negative) should elicit faster responses. For example, subjects with negative attitudes toward African Americans are expected to respond more quickly to the target word if a negative (e.g., "disgusting") target stimulus was preceded by a Black face. Several studies (e.g., Fazio et al. 1986, 1995) supported these theoretical expectations, thus providing evidence in favor of the automatic activation of attitudes. Interestingly, effects were also found for sequential priming tasks in which subjects were exposed to subliminal (i.e., unconscious) primes. For example, Greenwald et al. (1996) found that subliminal primes facilitate responses to semantically related targets. Taken together, at the end of the 1990s, the sequential priming paradigm was the dominant approach when it came to studying the automatic (i.e., implicit) activation of attitudes.

However, shortly before the turn of the millennium, research on implicit cognition witnessed another methodological innovation through the introduction of the implicit association test (IAT; Greenwald et al. 1998). Like the sequential priming procedure, IATs rest

on the assumption that the ease with which an individual can correctly sort stimuli (e.g., pictures/words) depends on the automatic (i.e., implicit) associations with the concepts of interest. In the classic IAT (Greenwald et al. 1998), participants complete a sequence of seven blocks of sorting tasks (see Table 1.1). After completing two practice blocks for concept and attribute discrimination, participants are instructed to press a left-key (e.g., "E") if the presented stimulus is either a positive adjective or pertaining to concept 1 and to press a right-key (e.g., "I") if the stimulus is either a negative adjective or pertaining to concept 2. After two such blocks, participants complete another two blocks in which the combinations are reversed (i.e., positive + concept 2 vs. negative + concept 1). Block order is counterbalanced across participants and stimuli are presented in random order.

Table 1.1: Sequence of Trial Blocks in the Classic IAT

Block	No.of Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Concept discrimination	Concept 1	Concept 2
2	20	Attribute discrimination	Positive words	Negative words
3	20	Practice combined block1	Positive + Concept 1	Negative + Concept 2
4	40	Combined block1	Positive + Concept 1	Negative + Concept 2
5	20	Reversed concept discrimination	Concept 2	Concept 1
6	20	Practice combined block 2	Positive + Concept 2	Negative + Concept 1
7	40	Combined block 2	Positive + Concept 2	Negative + Concept 1

Notes: The IAT score is based on data from blocks 3, 4, 6, and 7.

Finally, IAT scores are obtained by computing and averaging the differences in the mean response latencies between blocks 3 & 4 and blocks 6 & 7, also taking into account the variability of response latencies. In the classic IAT, the resulting IAT score can take on values between -2.0 and 2.0, thus providing a measure for both the direction and strength of a person's implicit attitude.

Due to its mostly satisfactory psychometric properties (e.g., predictive validity; see Greenwald et al. 2009), the IAT soon became the most widely used measure to capture implicit attitudes or stereotypes (see section 1.4). However, IATs have also attracted criticism (e.g., Blair 2002; Blanton et al. 2009). From a structural perspective, IATs were limited to providing relative assessments of target objects. "For instance, if a person shows a preference for candy bars over apples, we cannot tell whether he or she has a particular yen for Snickers, an aversion to Granny Smiths, or some combination of both." (Payne and Gawronski 2010, 8). To address this particular limitation, researchers have developed new (second generation) measures that allow for the measurement of implicit attitudes toward single target concepts. For example, the Go/No-Go Association Task (GNAT; Nosek and Banaji 2001), the Single-Category IAT (Karpinski and Steinman 2006), and the Single-Target IAT (Wigboldus et al. 2004) all share a concern for measuring implicit attitudes toward a single target object.

Much in the way that the IAT inspired researchers to develop similar but better (e.g., more reliable) IAT-related measures, priming paradigms were also refined in an attempt to further our understanding of semantic (e.g., Banaji and Hardin 1986) and affective (e.g., Wittenbrink et al. 1997) priming effects. For instance, Payne and colleagues (2005) developed the affect misattribution procedure (AMP), a sequential priming task that no longer relied on the measurement of response latencies. As such, the AMP made it possible to circumvent some of the major problems of the IAT (e.g., low reliability), resulting in high reliability and large effect sizes (for a meta-analysis, see Cameron et al. 2012). Finally, and more important for this dissertation, consumer psychologists identified the necessary conditions for subliminal advertising to be effective. Although considered a cargo-cult science for many years (e.g., Pratkanis 1992), recent studies (e.g., Bermeitinger et al. 2009; Karremans et al. 2006) suggest that subliminal ads may affect individuals' product choices, provided the ads represent goal-relevant cognitions and subjects are motivated to pursue the respective goals. Owing to the lack of conceptual replication of this paradigm in the political domain, I have evaluated its applicability in the context of political fear campaigns and right-wing authoritarianism (see chapter B1).

In sum, researchers of implicit social cognition have been very productive in terms of de-

veloping implicit attitude measures (for an overview, see Table 1.2). Thanks to these efforts, scholars can now draw from a large number of implicit measures that may help overcome the limitations (e.g., social desirability) of explicit (i.e., self-reported) measures. The availability of implicit measures has contributed to a very versatile and fast-growing field of research, providing some of social psychology's most intriguing insights.

Table 1.2: Overview of Implicit Measures

Implicit Measures			
1st Generation	Sequential Priming		
	Semantic Priming (Meyer and Schvaneveldt 1971)	Evaluative (Affective) Priming (Fazio et al. 1986)	IAT (Greenwald et al. 1998)
2nd Generation		Affect Misattribution Procedure (AMP) (Payne et al. 2005)	Single-Category IAT (SC-IAT) (Karpinski and Steinman 2006) Single-Target IAT (ST-IAT) (Wigboldus et al. 2004) Go/No-Go Association Task (GNAT) (Nosek and Banaji 2001) Extrinsic Affective Simon Task (De Houwer 2003) Single-Block IAT (Teige-Mocigemba et al. 2008) Brief IAT (Sriram and Greenwald 2009) Recoding-Free IAT (Rothermund et al. 2009)

1.4 Does Implicit Social (Political) Cognition Really Matter?

In recent years, implicit measures, most notably IATs, have been used for the assessment of implicit attitudes toward a very diverse range of concepts, such as smoking (Swanson et al. 2001), romantic partners (Banse and Kowalick 2007), and political parties (e.g., Friese et al. 2007). Unsurprisingly, the vast majority of research on implicit social cognition has focused on the (i) detection and (ii) impact of implicit stereotypes against African Americans. Using various implicit measures, researchers (e.g., Devine 1989; Fazio et al. 1995) have demonstrated that implicit racial stereotypes are widespread among White Americans. In addition, several studies have shown that implicit stereotypes against Black Americans can result in discriminative behavior. For instance, implicit stereotypes were found to negatively affect hiring decisions (Ziegert and Hanges 2005), medical treatment recommendations (Green et al. 2007), nonverbal behavior (Dovidio et al. 1997, 2002; Fazio et al. 1995), expectations of academic ability (Amodio and Devine 2006), and even police officers' decisions to shoot (Correll et al. 2002, 2007; Glaser and Knowles 2008). At this point, it should be noted that similar results were also found in different cultural contexts such as for Muslims in Sweden (Rooth 2007) or non-White job candidates in the UK (Beattie et al. 2013).

Of course, the investigation of implicit stereotypes has not been limited to racial stereotypes. In fact, implicit gender stereotypes have received a substantial amount of attention as well. For example, Nosek and colleagues (2002) found robust associations of "female" with "family" and "liberal arts", instead of "career" and "science", among visitors to a demonstration web site. Like implicit racial bias, implicit gender stereotypes have been held responsible for discrimination in the workplace (e.g., Rudman and Glick 2001), in academic science (e.g., Reuben et al. 2014), and in the legal profession (Levinson and Young 2010). However, as I will argue in the third paper (see chapter B3), research on the causal mechanism between female gender identity, implicit gender stereotypes, and career-relevant outcomes has been less convincing. For this reason, we have replicated previous experimental research (Moss-Racusin et al. 2012) on the detrimental effects of subtle gender bias using a more sophisticated experimental design.

Interestingly, in most of the above-mentioned studies, outcomes were not predicted by explicit measures. Accordingly, in their meta-analysis, Greenwald et al. (2009) concluded that IATs are better predictors of behavior in socially sensitive domains than self-reported attitudes are. Notwithstanding these results, research and researchers on implicit social cognition have also been the target of critique. Most prominently, Tetlock and Mitchell (2009, 6) have claimed

that "there is no evidence that the IAT reliably predicts class-wide discrimination on tangible outcomes in any setting." What lies at the heart of their argument is the call to move beyond the laboratory and conduct studies in "organizations, schools, hospitals, and other contexts in which implicit bias is of direct concern" (Oswald et al. 2015, 569). Other critiques (e.g., Blanton et al. 2009; Oswald et al. 2013, 2015) have blamed IATs for their arguably small effect sizes. While critics are certainly right in their call for more ecological validity, it would be careless to ignore the sheer overwhelming evidence supporting the argument that implicit measures predict certain behaviors better than self-reported attitudes do. In addition, IAT effects were found to be quite substantial for behaviors other than interracial behavior. Most importantly, Greenwald et al. (2009) found that IAT effects were largest for the prediction of political preferences. Therefore, contrary to the original caricature, implicit measures seemed to predict even highly deliberative and controlled behavior, such as voting for political parties and candidates. However, mostly because of a seminal study by Galdi et al. (2008), the dominant opinion still was that implicit attitudes would only predict the voting behavior of those unwilling or unable to report their voting intention. Hence, implicit measures were seen as a potential remedy to the challenge of undecided voters. Unfortunately, subsequent research could not replicate Galdi et al.'s (2008) results. If anything, implicit attitudes were found to be significant but even worse predictors – compared to explicit measures – of voting behavior among both decided and undecided voters (e.g., Friese et al. 2012) and to provide only little predictive validity over and above explicit measures. One major limitation of these studies, however, was that they were all conducted ahead of political elections, arguably a context in which attitudes are rather well elaborated, thus leaving little room for implicit measures to predict behavior beyond explicit measures. By contrast, implicit measures were expected to provide more incremental validity in the case of specific political issues, such as in the Galdi et al. (2008) study. Against the background of this debate, I conducted the first examination of the predictive validity of implicit attitudes in the context of federal votes in Switzerland. In my second paper (see chapter B2), I report on three studies in which implicit attitude measures were used to predict vote choice among both decided and undecided voters.

The research covered in this section by no means represents an exhaustive summary of the literature on implicit attitudes and stereotypes. Nevertheless, I have addressed some of the most intensely debated topics in the research on implicit social cognition and have illustrated how my research connects to these important agendas. In the next section, I briefly address the motivation of my research and the goals of my dissertation.

1.5 Motivation and Background of this Dissertation

The present dissertation is mainly the result of an observation, namely, that social psychologists have largely ignored Switzerland as a testing ground for implicit social (political) cognition. This neglect is most unfortunate. After all, the Swiss political system offers many opportunities to run studies on implicit political cognition that are impossible to execute in other countries. For instance, the direct-democratic process entails frequent voting on political issues. Citizens have the rights to propose changes to the constitution (i.e., launch a popular initiative) and to ask for a referendum on any law that has been passed by the parliament. Votes are held four times a year, and while most votes are on compulsory referendums, initiatives usually receive more attention from the media and public. For example, recent campaigns by the Swiss People's Party (SVP) have spurred a heated debate on the political culture in Switzerland, leading some (Vallely 2007) to describe Switzerland as "Europe's heart of darkness." In its campaigns, the SVP repeatedly attacked social out-groups such as immigrants or Muslims, even going as far as to successfully impose a ban on minarets. One major aspect of these campaigns was the use of highly threatening illustrations of immigration or Islam in poster ads. As such, SVP posters depict the type of societal threat that has been found to trigger authoritarianism even when the presentation of stimuli is subliminal (Feldman et al. 2010).

On the basis of these findings and drawing from recent insights into the working of subliminal advertising, I ran two experiments (see chapter B1) in which participants were subliminally primed with poster ads that were used by the SVP in their campaigns against immigrants and Muslims. My theoretical expectation was that the subliminal poster ads would only increase anti-immigrant or anti-Islam attitudes and behavior if subjects scored highly for right-wing authoritarianism (and thus were motivated to support political action against these groups).

In a second research endeavor (see chapter B2), I addressed the scientific debate on the predictive value of implicit attitude measures for the prediction of voting behavior. As already mentioned, previous research took place in the context of elections and thus in a setting in which implicit measures most likely overlapped with explicit attitudes. As a result, implicit attitudes will have a hard time predicting vote choice beyond explicit attitudes. In a different setting, however, such as the Swiss voting system, implicit measures may allow for higher predictive validity and be particularly useful for predicting the vote of undecided individuals. The latter represent a major challenge to anyone interested in predicting the outcome of an election or vote. As a matter of fact, the Swiss experience shows that even three weeks ahead

of a vote, an average 10 percent of respondents claim to be undecided in polls of the leading polling firm. Although this may seem a small number, undecided voters can be pivotal for the outcome of, in particular, close and highly contested votes. Therefore, political pollsters and political science in general would undoubtedly benefit from a better understanding of undecided voters' ultimate vote choice. For this reason, I conducted three studies ahead of votes for very diverse referendums, such as the implementation of a minimum wage, the purchase of fighter jets, the introduction of public health insurance, and a very controversial referendum on immigration.

Finally, in the third (co-authored) paper (see chapter B3), we have replicated previous research (Moss-Racusin et al. 2012) on the role of implicit or subtle gender bias in the domain of academic science. The particular motivation for this work did not stem from the Swiss experience or context but rather from the fact that new experimental designs are now available for the investigation of indirect effects. In particular, we have used Imai et al.'s (2011) parallel encouragement design (PED) to examine the causal mechanism between female gender, competence perceptions, and outcomes that are relevant to the pursuit of academic science careers. In addition, we have administered a gender-science IAT to examine whether implicit gender stereotypes moderate these indirect effects.

In sum, the goals of this dissertation are to set the stage for research on implicit political cognition in Switzerland and to further our understanding of implicit processes and measures in the political and social domain. As such, the three papers that constitute this dissertation do not follow a red thread but rather address some of the cutting-edge topics in implicit social cognition research. I shall briefly summarize and discuss the main findings of my dissertation in the next section.

1.6 Results and Discussion

In the first paper (see chapter B1), I found that subliminal versions of poster ads that were used in right-populist fear campaigns in Switzerland did indeed affect those who, as demonstrated in the literature on right-wing authoritarianism, support policies against unconventional groups. In both a laboratory (Study 1) and web experiment (Study 2), authoritarian participants were more likely to sign a petition against foreign students (Study 1) or support a ban on mosques and minarets (Study 2) if exposed to subliminal poster ads. After comparing the group means between high authoritarians in the control and prime conditions, I found weak (Study 1: Cohen's $d = 0.26$) and moderate effect sizes (Study 2: $d = 0.39$). By

contrast, there were no effects among low-authoritarian participants. Thus, consistent with recent research on subliminal advertising, subliminal cues may influence individuals' attitudes and behavior provided the primes are relevant to their motivational goals. However, effects were considerably smaller than in previous research, most likely as a result of the different motivational goals that were studied. Of course, physical needs (e.g., quenching one's thirst) represent more fundamental motivations than do political goals (e.g., curb immigration). Nevertheless, my results imply that political parties may have a tool to secretly persuade those who already endorse the main goals and political beliefs of the party.

In the second paper (see chapter B2), I consistently found that explicit (i.e., self-reported) attitudes were better predictors for decided voters than were implicit attitudes obtained from Single-Target IATs. However, among undecided voters, implicit and explicit attitudes were equally good predictors, and implicit attitudes predicted voting behavior descriptively, but not significantly, better for undecided voters. Finally, explicit attitudes predicted voting descriptively (but, again, not significantly) better for decided voters. Taken together, results suggest that the utility of implicit attitudes may be higher in the context of voting on political issues than in elections, but still not as high as initially believed or hoped-for.

Finally, in the third paper (see chapter B3), we found only weak implicit gender stereotypes (mean IAT score = 0.20) among a sample of more than 1,400 MTurk subjects. Unlike the replicated research, we did not find negative effects of female gender on perceived hireability for a lab manager position or for salary conferral. Rather, we found that participants would offer a higher starting salary to female students, not because they perceived them as more competent but because they viewed them as more likeable than male student applicants. Interestingly, the indirect effect between student gender, likeability, and salary offer was not moderated by subjects' implicit gender stereotypes obtained from a gender-science IAT. Since our sample was fairly liberal and well educated, we conclude that future cohorts of science faculty may be less likely to exhibit implicit gender stereotypes and thus less likely to discriminate against female students.

To sum up, my results suggest that implicit social cognition does indeed matter, but only under very specific conditions and sometimes not as much as purported in the literature. For instance, the effectiveness of subliminal ads depends on individuals' motivational goals and the goal-relevance of the prime. In a similar vein, the value of implicit attitudes for the prediction of voting behavior is highly contingent on the degree of cognitive elaboration of attitudes and, most likely, other moderators. Finally, implicit gender stereotypes may not be as pervasive as has been argued in previous research on the topic.

My findings raise important questions that could be addressed in future research. For example, using different stimuli and psychological constructs (other than right-wing authoritarianism), can the subliminal advertising paradigm be replicated? Moreover, what about the temporal persistence of the effects? Given the very diverse and complex issues that Swiss citizens typically vote on, would we still find low incremental validity of implicit attitudes or would they be more useful as a result of poorly elaborated attitudes? Finally, can the PED that was used in the third paper be applied to other domains and target populations to study the causal mechanisms between implicit stereotypes and discrimination?

References

Amodio, D.M., and Patricia G. Devine. 2006. Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior. *Journal of Personality and Social Psychology*, 91(4), 652-661.

Banaji, M.R., and Curtis D. Hardin. 1996. Automatic stereotyping. *Psychological Science*, 7, 136-141.

Banse, R., and Claudia Kowalick. 2007. Implicit attitudes towards romantic partners predict well-being in stressful life conditions: Evidence from the antenatal maternity ward. *International Journal of Psychology*, 42(3), 149-157.

Beattie, G., Cohen, D., and Laura McGuire. 2013. An Exploration of Possible Unconscious Ethnic Biases in Higher Education: The Role of Implicit Attitudes on Selection for University Posts. *Semiotica*, 197, 171-201.

Bermeitinger, C., Goelz, R., Johr, N., Neumann, M., Ecker, U.K.H., and Robert Doerr. 2009. The hidden persuaders break into the tired brain *Journal of Experimental Social Psychology*, 45(2), 320-326.

Blair, I.V. 2002. The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, 6, 242-261.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., and Gregory Mitchell. 2009. *Journal of Applied Psychology*, 94(3), 567-582.

Bruner, J.S. 1957. On Perceptual Readiness. *Psychological Review*, 64, 123-152.

Cameron, C.D., Brown-Iannuzzi, J., and Brian K. Payne. 2012. Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, 16, 330-350.

Correll, J., Park, B., Judd, C.M., and Bernd Wittenbrink. 2002. The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology*, 83(6), 1314-1329.

Correll, J., Park, B., Judd, C.M., Wittenbrink, B., Sadler, M.S., and Tracie Keesee. 2007. Across the Thin Blue Line: Police Officers and the Racial Bias in the Decision to Shoot. *Journal of Personality and Social Psychology*, 92(6), 1006-1023.

De Houwer, J. 2003. The Extrinsic Affective Simon Task. *Experimental Psychology*, 50, 77-85.

Devine, P.G. 1989. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.

Dovidio, J.F., Evans, N., and Richard B. Tyler. 1986. Racial Stereotypes: The Contents of their Cognitive Representations. *Journal of Experimental Social Psychology*, 22, 22-37.

Dovidio, J.F., Kawakami, K., Johnson, C., Johnson, B., and Adaiah Howard. 1997. On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, 33(5), 510-540.

Dovidio, J.F., Kawakami, K., and Samuel L. Gaertner. 2002. Implicit and Explicit Prejudice and Interracial Interaction. *Journal of Personality and Social Psychology*, 82(1), 62-68.

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., and Frank R. Kardes. 1986. On The Automatic Activation of Attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.

Fazio, R.H., Jackson, J.R., Dunton, B.C., and Carol J. Williams. 1995. Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline? *Journal of Personality and Social Psychology*, 69, 1013-1027.

Feldman, S., Lavine, H., Lodge, M., and Bradley Verhulst. 2010. Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli. Paper presented at the annual meeting of the International Society of Political Psychology, San Francisco, CA.

Friese, M., Bluemke, M., and Michaela Wänke. 2007. Predicting Voting Behavior with Implicit Attitude Measures. *Experimental Psychology*, 54(4), 247-255.

Friese, M., Smith, C.T., Plischke, T., Bluemke, M., and Brian A. Nosek. 2012. Do implicit attitudes predict voting behavior particularly for undecided voters? *PLoS ONE*, 7(8).

Gaertner, S.L., and John P. McLaughlin. 1983. Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics. *Social Psychology Quarterly*, 46, 23-30.

Galdi, S., Arcuri, L., and Bertram Gawronski. 2008. Automatic Mental Associations Predict Future Choices of Undecided Decision-Makers. *Science*, 321, 1100-1102.

Gawronski, B., Hofmann, W., and Christopher J. Wilbur. 2006. Are "implicit" attitudes unconscious? *Consciousness and Cognition*, 15(3), 485-499.

Glaser, J., and Eric D. Knowles. 2008. Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, 44, 164-172.

Green, A.R., Carney, D.R., Pallin, D.J., Ngo, L.H., Raymond, K.L., Iezzoni, L.I., and Mahzarin R. Banaji. 2007. Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients. *Journal of General Internal Medicine*, 22(9), 1231-1238.

Greenwald, A.G., and Mahzarin R. Banaji. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27.

Greenwald, Anthony G., Draine, S.C., and Richard L. Abrams. 1996. Three Cognitive Markers of Unconscious Semantic Activation. *Science*, 273, 1699-1702.

Greenwald, A.G., McGhee, D.E., and Jordan L.K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., and Mahzarin R. Banaji. 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.

Hahn, A., Judd, C.M., Hirsh, H.K., and Irene V. Blair. 2014. Awareness of Implicit Attitudes. *Journal of Experimental Psychology: General*, 143(3), 1369-1392.

Helmholtz, H. 1925. Treatise on physiological optics (3rd ed., Vol. 3). Menasha: Banta. (Original 3rd ed. published in 1910).

Imai, K., Keele, L., Tingley, D., and Teppei Yamamoto. 2011. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, 105(4), 765-789.

Karpinski, A., and Ross B. Steinman. 2006. The single category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, 91, 16-32.

Karremans, J.C., Stroebe, W., and Jasper Claus. 2006. Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. *Journal of Experimental Social Psychology*, 42, 792-798.

Levinson, J.D., and Danielle Young. 2010. Implicit Gender Bias in the Legal Profession: An Empirical Study. *Duke Journal of Gender Law & Policy*, 18(1), 1-33.

Meyer, D.E., and Roger W. Schvanefeldt. 1971. Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations. *Journal of Experimental Psychology*, 90, 227-234.

Meyer, D.E., and Roger W. Schvanefeldt. 1976. Meaning, Memory Structure, and Mental Processes. *Science*, 192(4234), 27-33.

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., and Jo Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences PNAS*, 109(41), 16474-16479.

Neely, J.H. 1977. Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention. *Journal of Experimental Psychology: General*, 106, 226-254.

Nosek, B.A., and Mahzarin R. Banaji. 2001. The Go/No-Go Association Task. *Social Cognition*, 19, 625-666.

Nosek, B.A., Banaji, M.R., and Anthony G. Greenwald. 2002. Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site. *Group Dynamics*, 6(1), 101-115.

Oswald, F.L., Mitchell, G., Blanton, H., Jaccard, J., and Philip E. Tetlock. 2013. Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies. *Journal of Personality and Social Psychology*, 105(2), 171-192.

Oswald, F.L., Mitchell, G., Blanton, H., Jaccard, J., and Philip E. Tetlock. 2015. Using the IAT to Predict Ethnic and Racial Discrimination: Small Effect Sizes of Unknown Societal Significance. *Journal of Personality and Social Psychology*, 108(4), 562-571.

Payne, B.K., Cheng, C.M., Govorun, O., and Brandon D. Stewart. 2005. An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293.

Payne, B.K., and Bertram Gawronski. 2010. A History of Implicit Social Cognition: Where Is It Coming From? Where Is It Now? Where Is It Going? In: Gawronski, B., and Brian K. Payne (Eds.), *Handbook of Implicit Social Cognition*, 1-15. New York: The Guilford Press.

Pratkanis, Anthony. 1992. The cargo-cult science of subliminal persuasion. *Skeptical Inquirer*, 16, 260-272.

Reuben, E., Sapienza, P., and Luigi Zingales. 2014. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences PNAS*, 111(12), 4403-4408.

Rooth, Dan-Olof. 2007. Implicit Discrimination in Hiring: Real World Evidence (Discussion Paper No. 2764). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit / Institute for the Study of Labor.

Rosch E. 1975. Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104, 192-233.

Rothermund, K., Teige-Mocigemba, S., Gast, A., and Dirk Wentura. 2009. Minimizing the influence of recoding in the IAT: The Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology*, 62, 84-98.

Rudman, L.A., and Peter Glick. 2001. Prescriptive Gender Stereotypes and Backlash Toward Agentic Women. *Journal of Social Issues*, 57(4), 743-762.

Sriram, N., and Anthony G. Greenwald. 2009. The Brief Implicit Association Test. *Experimental Psychology*, 56(4), 283-294.

Swanson, J.E., Rudman, L.A., and Anthony G. Greenwald. 2001. Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion*, 15, 207-230.

Teige-Mocigemba, S., Klauer, K.C., and Klaus Rothermund. 2008. Minimizing Method-Specific Variance in the IAT: A Single-Block IAT. *European Journal of Psychological Assessment*, 24, 237-245.

Tetlock, P.E., and Gregory Mitchell. 2009. Implicit Bias and Accountability Systems: What Must Organizations Do to Prevent Discrimination? *Research in Organizational Behavior*, 29, 3-38.

Vallely, P. 2007. Switzerland, Europe's heart of darkness? Reprinted in the *Belfast Telegraph*, September 7, 2007. Available online: http://www.belfasttelegraph.co.uk/news/world-news/switzerland-europes-heart-of-darkness-28062699.html.

Wigboldus, D.H.J., Holland, R.W., and Ad van Knippenberg. 2004. Single target implicit associations. Unpublished manuscript.

Wittenbrink, B., Judd, C.M., and Bernadette Park. 1997. Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262-274.

Ziegert, J.C., and Paul J. Hanges. 2005. Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias. *Journal of Applied Psychology*, 90(3), 553-562.

Part B

Articles

1

Subliminal Fear Ads and Right-wing Authoritarianism: Experimental Evidence from Switzerland and the US

Livio Raccuia

RECENT advances in subliminal advertising research suggest that subliminal cues can affect behavior if two necessary conditions are satisfied. First, individuals need to be motivated to pursue a certain goal to which, secondly, the subliminally primed stimulus is relevant. This article argues that both conditions are satisfied in the case of high authoritarian individuals if they are exposed to subliminal presentations of threatening political poster ads. Drawing from research on authoritarianism, an interaction between political poster ads that pertain to societal threats and right-wing authoritarianism (RWA) was tested in a laboratory and web experiment using Amazon's Mechanical Turk (MTurk). Using two different primes and priming procedures, both main and interaction effects of subliminal versions of right-populist poster ads were found. Given that these poster ads were actually used in Swiss referendum campaigns, the implications of these findings for the use in political campaigns are discussed.

1.1 Introduction

After years of doubt and skepticism about the integrity of its findings, subliminal advertising research has recently experienced substantial development through the efforts of various scholars. Arguing that previous studies on the matter had largely ignored the necessary conditions for subliminal advertising to be effective, scholars were able to identify two such conditions. First, individuals need to be motivated to pursue a certain goal. Second, the primed stimulus needs to be a goal-relevant cognition. Conclusions were drawn from studies in which thirsty (i.e., motivated to quench their thirst) individuals' drinking behavior and brand choice were found to be affected by subliminal cues such as "thirst" (Strahan et al. 2002) or "Lipton Ice" (Karremans et al. 2006; Verwijmeren et al. 2011) much like tired people (i.e., those motivated to enhance concentration) were affected in their choice of energy pills by subliminal presentations of brand logos (Bermeitinger et al. 2009). Interestingly, however, no attempts have been made to apply the aforementioned paradigm to the domain of political advertising. In this article, I seek to fill this void by conceptually replicating this paradigm in the domain of political fear campaigns. While fear ads have received much attention in the study of American politics (e.g., Brader 2006), their effects have not been investigated in the context of the newly developed paradigm of subliminal advertising. Given the threatening nature of fear ads and the necessary focus on people's motivations, I build my theoretical model on the debate on authoritarianism under threat. Originally inspired by Adorno et al.'s (1950) attempt to explain mass support for nazism, research on authoritarianism later shifted its focus to the role of perceived threat in the activation of authoritarianism. While researchers initially observed a simultaneous occurrence of threats and increased authoritarian attitudes among individuals (Doty et al. 1991), more recent research has confined the activation of authoritarian predispositions to forms of perceived political and societal threats (e.g., Feldman and Stenner 1997). That is, rather than threats to personal well-being, high authoritarians were found to react sensitively if they perceived threats to the political and social order. As observed by several authors (e.g., Arendt et al. 2015; Betz 2013) right-populist parties in Europe make frequent use of threat appeals in their political advertising in general and poster campaigns in particular. For example, the Swiss People's Party (SVP) has used threat appeals in a number of poster campaigns. Intended to induce a sense of threat and unease among citizens, these poster ads led to public outcry each time the party won a campaign. Unsurprisingly, the debate centered on whether political success can be explained by such fear campaigns.[1]

In this article, I address this debate and add to both the literature on authoritarianism under threat and the debate on the impact of political fear campaigns. Building on the current view in research on authoritarianism, I argue that the type of stimulus typically used in political fear campaigns can affect the high authoritarian component of the electorate in a direction that is intended by the message and thus the political actor behind the campaign (e.g., the

[1] Throughout this article, the term "fear campaign" will only refer to right-populist campaigns.

political party). Therefore, I hypothesize an interaction between threat appeals in the form of political poster ads and right-wing authoritarianism (RWA). Moreover, I expect to find this interaction even when exposure to threat appeals is subliminal. In fact, Feldman et al. (2010) recently reported evidence in support of this effect. In other words, even though their subjects were not consciously aware of the stimuli, they were still influenced by the primes provided they scored high on authoritarianism. In light of these findings, I tested the aforementioned hypothesis using subliminal stimuli. Unlike supraliminal priming, where individuals are consciously aware of the primed stimulus, subliminal priming allows researchers to uncover the effects of poster ads free of the "damping effect of negation" (Arendt 2013). In fact, a large body of literature (e.g., Gawronski and Bodenhausen 2006; Wegener and Petty 1997) holds that individuals may be very critical toward the information they receive if they understand it is an attempt to persuade them. This may eventually lead to the rejection of the information and thus the absence of an effect of the primed stimulus. However, the unobtrusive nature of subliminal priming prevents such correction processes (Wegener and Petty 1997) and thus simplifies the analysis plan by omitting the potential moderating effect of negation. By conducting both a laboratory and online experiment, I address concerns about external validity. Furthermore, I examine the robustness of the findings by using different primes and priming procedures (i.e., the movie and priming task). The hope is that the present research will for the first time provide empirical evidence on the efficacy of political fear campaigns while at the same time offering insight about the workings of subliminal advertising.

Data for this article were collected in one laboratory and one web experiment. Both were written and administered using the Inquisit software and the web experiment was hosted at http://www.millisecond.com. For the online experiment, the subjects were recruited and paid through Amazon's Mechanical Turk (MTurk, http://www.mturk.com) system. In total, 633 Swiss and non-Hispanic White adult Americans participated in two 2 (treatment vs. control) x 2 (low vs. high authoritarians) between-subject designs.

In the remainder of the article, I will first briefly discuss both authoritarianism and subliminal advertising research. In doing so, I will argue that political fear campaigns allow for the application of the subliminal advertising paradigm because they satisfy its necessary conditions when applied to high authoritarian individuals. Finally, I will present and discuss evidence from two experiments in which subjects were exposed to subliminal political poster ads that were used in right-populist campaigns against immigrants and Muslims in Switzerland.

1.2 Literature Review and Theory

1.2.1 Subliminal advertising

Unsurprisingly, the practice of influencing people's preferences and behavior without them being consciously aware of it has received much attention in advertising research where consumer choices and profit are at stake. Indeed, the very origins of subliminal priming can be

traced back to James Vicary's claim that sales of popcorn and coke dramatically increased when the words "eat popcorn" and "drink Coke" were secretly flashed on the screen in a New Jersey movie theatre. Although his experiment was later revealed as a publicity hoax (Pratkanis 1992), the concept of subliminal advertising has kept researchers intrigued across disciplines. In regards to political psychology, subliminal priming was mostly used in order to demonstrate the automaticity of affect (e.g., Lodge and Taber 2005) for political leaders, groups, and issues but less so for studying the impacts of subliminal versions of political ads. This may, in large part, be explained by the fact that evidence in favor of the effects of subliminal advertising was rather poor for many decades following its first mention. For example, a meta-analysis by Trappey (1996, 517) revealed that only "very little effect" on consumer choice was found in nine studies (Byrne 1959; Caccavale et al. 1981; Champion and Turner 1959; Cuperfain and Clarke 1985; Gable et al. 1987; George and Jennings 1975; Hawkins 1970; Kilbourne et al. 1985; Weinstein et al. 1986)

The emergence of the subliminal priming paradigm

After decades of contradicting evidence on subliminal advertising, progress was eventually made through the efforts of scholars who claimed that prior research on the matter had simply overlooked the necessary conditions of subliminal advertising. According to these authors, one such condition was individuals' motivation to pursue a goal to which the prime is relevant. For example, in both, Strahan et al. (2002) and Karremans et al. (2006), subliminally priming subjects with thirst-related cognitions (e.g., "thirst") or brand names (e.g., "Lipton Ice") only resulted in positive effects if the individuals were already thirsty. Researchers concluded that being thirsty, and thus motivated to quench one's thirst, was a necessary precondition for thirst-related subliminal cues to be effective. The second necessary condition follows from the first. If individuals pursue a certain goal, then subliminal stimuli may affect their strategy or method used to accomplish the goal provided the stimuli are goal-relevant by either generally referring to the goal (e.g., "thirst") or promoting a means to actually attain the goal (e.g., "Lipton Ice"). Whereas the first form is commonly referred to as "goal priming" and accounts for the majority of studies in subliminal priming research, the latter form of actually priming a brand name is characteristic of subliminal advertising. In sum, individuals' motivation to pursue a certain goal and the goal relevance of the prime were identified as necessary conditions of subliminal advertising.

Interestingly, the above paradigm has found support in several conceptual replications. For example, Bermeitinger et al. (2009) adapted Strahan et al.'s (2002) and Karremans et al.'s (2006) concept to the domain of concentration. In line with their expectations, they found the product choice of tired (i.e., motivated) subjects to be affected by subliminal ads of virtual brands of dextrose pills. Taken together, these studies provide evidence for the effectiveness of subliminal advertising in mental (e.g., concentration) and physical (e.g., thirst) states. To my knowledge, however, the presented paradigm has not yet been applied to the domain of political advertising. Although both Stewart and Schubert (2006) and Weinberger

and Westen (2008) dedicated their attention to testing the efficacy of the "RATS" prime[2], they did not build their model on subjects' motivations and the goal relevance of the prime. The sole, yet only approximate, replication of the paradigm was conducted by Hassin et al. (2007, 2009). In a series of experiments, Hassin and colleagues reported significant effects of subliminal exposure to the Israeli (2007), and Italian (2009) flag on a number of attitudinal and behavioral outcomes. Given that the effects were moderated by participants' identification with Israeli or Italian nationalism (IWIN), they concluded that individuals' motivations are key in understanding the effects of subliminal cues.

1.2.2 Authoritarianism (under threat)

> *"Threat appears to be critical to the activation of authoritarianism."*
> (Feldman and Stenner 1997, 765)

Ever since Adorno et al.'s (1950) groundbreaking work on the authoritarian personality, the role of threat has been at the epicenter of research dealing with authoritarianism. Yet although there is a long history of research on the relationship between authoritarianism and threat using aggregate data (e.g., Fromm 1941; Rokeach 1960; Sales 1972; Wilson 1973), research has only recently shifted its focus to individual-level data. Inspired by Doty et al.'s (1991) finding that authoritarianism becomes more pronounced in US citizens during periods of societal threat – and subsequently recedes when that threat is over – Feldman and Stenner (1997) were the first to provide evidence for this interaction using individual-level data. After testing the interaction of authoritarianism and various types of threat (e.g., political, economic, fear of war), they found intolerant and punitive attitudes toward minority groups to be significantly predicted by an interaction of authoritarianism and perceived societal threat. The same, however, could not be said for threats to personal well-being. Hence, Feldman and Stenner (1997, 765) concluded that "[a]uthoritarianism is activated when there is a perception that the political or social order is threatened". Note that, according to the authors, authoritarianism is activated rather than increased upon exposure to a societal threat. In fact, the authors argue that societal threats simply activate authoritarian predispositions and thus lead to more pronounced manifestations of authoritarianism under conditions of perceived threat.

In light of this evidence, it has become the dominant approach in authoritarianism research to regard authoritarianism as a latent predisposition that comes to the fore in the presence of normative threat (e.g., Duckitt 1989; Stenner 2005). Until very recently, however, threat exposure has been limited to studies of explicit threat stimuli. That is, individuals were always aware of the threatening stimuli they were exposed to. Adding to this, Feldman et al. (2010) examined whether high authoritarians also respond to threatening stimuli if the latter are not consciously perceived. In their experiment, Feldman et al. (2010) subliminally exposed

[2]The so-called "RATS" prime was allegedly embedded in a Republican TV ad that was aired during the 2000 US presidential election as an attack on Democratic candidate Al Gore.

individuals to different types (general, mortality, normative/group) of threatening (negative) words along with pseudo-words. In line with Feldman and Stenner's previous (1997) research, they found a significantly higher probability of the correct identification of threatening words, which was moderated by the subjects' level of authoritarianism. Importantly, however, the interaction was only significant for words referring to threats to normative order and group cohesion (e.g., "foreign", "they", and "diverse") but not so for general words (e.g., "pain") and words pertaining to mortality (e.g., "death"). Therefore, it seems that high authoritarians are especially sensitive to threat-related messages to in-group norms and values, even if such messages are operating outside of their conscious awareness.

1.3 Theoretical Model

In this article, I argue that authoritarianism under threat and political fear campaigns together provides an adequate domain for a conceptual replication of the subliminal advertising paradigm. My argument rests on the notion that while authoritarianism satisfies the first necessary condition of subliminal advertising (i.e., motivation), political fear campaigns and their corresponding poster stimuli do so for the second (i.e., goal relevance). Consider first the motivations of individuals. In a number of studies, high authoritarians have been characterized as individuals who are motivated to defend social control, stability, and cohesion (e.g., Duckitt 2006). As a result of their underlying motivational goals, high authoritarians typically exhibit distinct prejudice toward unconventional groups, such as drug dealers (Duckitt 2006), rock stars (Duckitt 2006), homosexuals (Butler 2000), obese people (Butler 2000), and immigrants (e.g., Cohrs et al. 2005). In fact, immigrants attract a great deal of high authoritarians' attention and prejudice, mainly because of high authoritarians' chronic concern about the impact of immigrants on in-group norms and values. For example, the studies of both Manganelli Rattazzi et al. (2007) and Cohrs et al. (2005) report highly significant correlations between right-wing authoritarianism (RWA) and prejudice toward islamic immigrants. Similar results were found for Switzerland where RWA significantly predicted support for persecution of immigrants who are unwilling to assimilate into the host culture (Thomsen et al. 2008). Thus, when it comes to politics, high authoritarians are particularly inclined to support political intolerance and even punitive solutions (Peterson et al. 1993). I therefore expect them to be especially prone to supporting political campaigns promoting anti-immigrant or anti-Islam policies.

1.4 Study 1

1.4.1 Participants and design

A total of 155 undergraduate and graduate students (80 female, 75 male) from the University of Zurich participated in exchange for 40 Swiss Francs. Their mean age was 25.7 years. When arriving in the lab, subjects were randomly assigned to either the treatment (prime) or control

(no prime) condition and given their personal RWA score in the form of a subject identity code (i.e., the subjects were unaware of the true nature of their code). Subjects' RWA scores had been previously collected in an online survey. The study was a 2 (treatment vs. control) x 2 (low authoritarians vs. high authoritarians) between-subjects factorial design.

1.4.2 Procedure

Upon their arrival in the lab, participants were told that the study was about their attitudes toward university policy. More precisely, financial and capacity problems at Swiss universities served as my cover story. The experiment then consisted of four parts. First, participants were asked to enter their subject identity code (i.e., RWA score) along with socio-demographic data. Second, they viewed a short movie (approx. 6 min.) about Swiss universities. The movied had been compiled from various television shows that addressed the increasing number of students at Swiss universities; subliminal primes were embedded in the movie. Concretely, in each condition 21 single frames were interspersed into the movie for a duration of 33 milliseconds. A movie was used instead of a priming task in a procedure that was in line with Cooper and Cooper (2002). The decision to choose a movie can be justified by two reasons. First, if placed at scene changes, the target primes require no backwards masking[3] due to the natural flicker of the scene change. Second, presenting a movie in accordance with the cover story improves the coherence and credibility of the study. While watching the movie, participants sat approximately 45 cm from a Dell computer with a screen resolution of 1400 x 900 pixels and a refresh rate of 60 Hz. The subliminal picture was placed at eight peripheral spots on the screen and appeared in an irregular order at one of those eight positions. Third, subjects were asked a series of questions about their attitudes on university policy (e.g., support for higher/lower tuition fees). Questions pertaining to immigration in general and to foreign students in particular were included in the questionnaire (see below). Finally, a two-step procedure similar to the funneled debriefing procedure proposed by Bargh and Chartrand (2000) was used in order to test whether the presentation of the primes was indeed subliminal. To this end, participants were first asked if they had noticed anything peculiar in the movie and were then informed that a picture had been placed subliminally. In addition, they were subjected to a forced-choice test in which they had to choose from a list of five primes (see Figure 1.1) the one they thought had been primed. After the subliminality check, participants were paid and thanked for their participation.

[3]Backwards masking means the subsequent masking of the prime by some other visual stimulus. This process wipes the residual activation for the target prime from the individuals' retinas, thereby establishing unconscious perception of the target stimulus.

Figure 1.1: Posters Used in Forced-Choice Test of Study 1

1.4.3 Independent variables

Prime condition

The Swiss People's Party's (SVP) campaign poster against immigration (Poster D in Figure 1.1) served as the treatment prime. The poster was used in the party's campaign against immigration between 2012 and 2014. It was chosen for three theoretical reasons. First, the poster displays a threat to Swiss society by showing black boots stomping on the Swiss flag. Hence, given the threatening character of the poster, it may well activate authoritarian predispositions among high authoritarian individuals. Second, the initiative and thus the corresponding poster suggest the implementation of immigration quotas. As such, the initiative can be seen as a means to reduce the influence of an out-group (i.e., immigrants) that high authoritarians are deeply concerned about. Therefore, as already explained in the previous section, the initiative – and thus the poster – meet the necessary condition of goal-relevance. Third, when the study was conducted, the poster campaign had already been in place for more than one year. I thus assumed that participants were familiar with the prime and that the message could thus be recognized and processed by the subjects. In total, 21 single frames of the prime were interspersed into the movie. In the control condition, subjects viewed the same movie with single frames of 33-ms duration that occurred at precisely the same points as in the treatment condition. However, the control prime was blank (i.e., merely white space), as in Cooper and Cooper (2002).

Authoritarianism

Prior to the lab sessions, the participants completed an online survey in which their level of authoritarianism was measured using the 12-item RWA-scale proposed by Funke (2005). The scale comprises 12 items from Altemeyer's (1996) RWA scale in a balanced way (i.e., dimension (3) x wording direction (2) x items (2)). Participants reported their level of agreement with these items using a 7-point Likert scale (1(7) = *strongly disagree* to 7(1) = *strongly agree*) (Cronbach's $\alpha = 0.79$). Global scores on the scale ranged from 12 to 84. For some of the analyses, the sample was split into low (LA) and high authoritarians (HA) based on the sample median (= 31). This median split resulted in 82 LAs and 73 HAs.

1.4.4 Dependent variables

Attitudes toward immigration
Participants' attitudes toward immigration were measured with two items ("Should more or fewer immigrants be allowed to live permanently in Switzerland?" [1 = *a lot less*, 5 = *a lot more*], "Do you think that immigrants enrich or endanger Swiss society?" [1 = *very much endanger*, 5 = *very much enrich*]). The two items were averaged to form an index for participants' attitudes toward immigration (Cronbach's $\alpha = 0.66$). Given the poster prime and its negative reference to "mass immigration," I expected a negative interaction between the treatment stimulus and authoritarianism.

Petition for foreign student quota
In addition, subjects were given the chance to communicate their position on educational immigration to the head of the University by signing a petition that required them to enter their first and last name as well as their place of residence. The petition called upon the principal to consider the implementation of a foreign student quota, and I expected the treatment to increase the probability that high authoritarian subjects would sign the petition.

1.4.5 Results and discussion

Subliminality check
Only one participant correctly identified the poster when asked to report any unusual things in the movie. In addition, one participant mentioned another SVP poster that had been used in 2007. For these two subjects, the priming intensity was above subliminality, and they were thus excluded from subsequent analyses. In the forced-choice test, 42% of participants in the treatment condition and 39% of participants in the control condition chose the correct prime. As this difference was not significant ($t(151) = -0.40$, $p = .34$) I concluded that subjects did not consciously perceive the prime.

Attitudes toward immigration
The descriptive statistics and correlations between authoritarianism and the dependent variables are presented in Table 1.1. As I expected, authoritarianism was strongly predictive of lower scores on the scale for immigration attitudes. I first tested whether subliminal exposure to the poster prime negatively affected subjects' attitudes toward immigration, and whether this effect would be moderated by authoritarianism. Participants' attitudes toward immigration were regressed onto the priming condition (1 = *treatment prime*, 0 = *control prime*), authoritarianism (standardized score), and their interaction. Interestingly, only authoritarianism exerted a significant negative effect ($\beta = -.39$, $t(152) = -3.55$, $p < .01$) on participants' attitudes toward immigration. As can be seen in Table 1.2, immigration attitudes were significantly ($t(151) = 4.95$, $p < .001$) lower in the high authoritarian condition than among low authoritarian subjects. However, both the priming condition and the interaction between the

priming condition and authoritarianism did not affect subjects' immigration attitudes.

Table 1.1: Descriptive statistics and correlations between right-wing authoritarianism and the dependent variables (Study 1)

	M	**SD**	**RWA**	**Immigration attitudes**	**Petition**
RWA	32.24	9.94	-		
Immigration attitudes	3.61	0.79	-0.49***	-	
Petition	0.24	0.43	0.21**	-0.20*	-

* p<.05, ** p<.01, *** p<.001.

Table 1.2: Dependent variables as a function of priming condition and authoritarianism (Study 1)

	Right-wing authoritarianism				**Priming condition**			
	Low (n=81)		**High (n=72)**		**Control (n=82)**		**Prime (n=71)**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Immigration attitudes	3.89	0.67	3.30	0.80	3.51	0.83	3.73	0.74
Petition	0.17	0.38	0.31	0.46	0.22	0.42	0.25	0.44

Petition for foreign student quota

In line with my expectation, right-wing authoritarianism was positively correlated (see Table 1.1) with the propensity to sign the petition for a foreign student quota. To analyze whether the subliminal poster ad affected subjects' propensity to sign the petition for a foreign student quota, I regressed signature of the petition (1 = *Yes*, 0 = *No*; in this manner, the dependent variable represents the percentage of participants signing the petition) onto the priming condition, authoritarianism, and their interaction. This analysis yielded no significant main or interaction effect of the prime on the likelihood of signing the petition. However, separate regression analyses for the effect of authoritarianism on the probability to sign the petition for subjects in the treatment and control condition revealed a positive effect in the treatment condition ($\beta = .13$, $t(70) = 2.84$, $p < .01$) but no effect in the control condition ($\beta = .06$, $t(81) = 1.30$, $p = .20$).[4] Furthermore, consistent with my theoretical expectation, the average support of the petition was significantly ($t(28) = 1.35$, $p = .09$) higher among high authoritarians in the treatment condition ($M = 0.38$) than among high authoritarians in the control condition ($M = 0.26$) (see Figure 1.2, right panel). According to Cohen (1988), the difference in the group means implies a small effect ($d = 0.26$) of the subliminal prime in the high authoritarian subgroup and no effect ($d = 0.03$) among low authoritarian subjects.[5]

[4] Logistic regression analyses yielded the same results.

[5] Cohen's (1988) standardized d statistic allows comparisons of effects across studies with different measurement units. Effects are classified as small if $d \approx 0.20$, medium if $d \approx 0.50$, and large if $d \approx 0.80$.

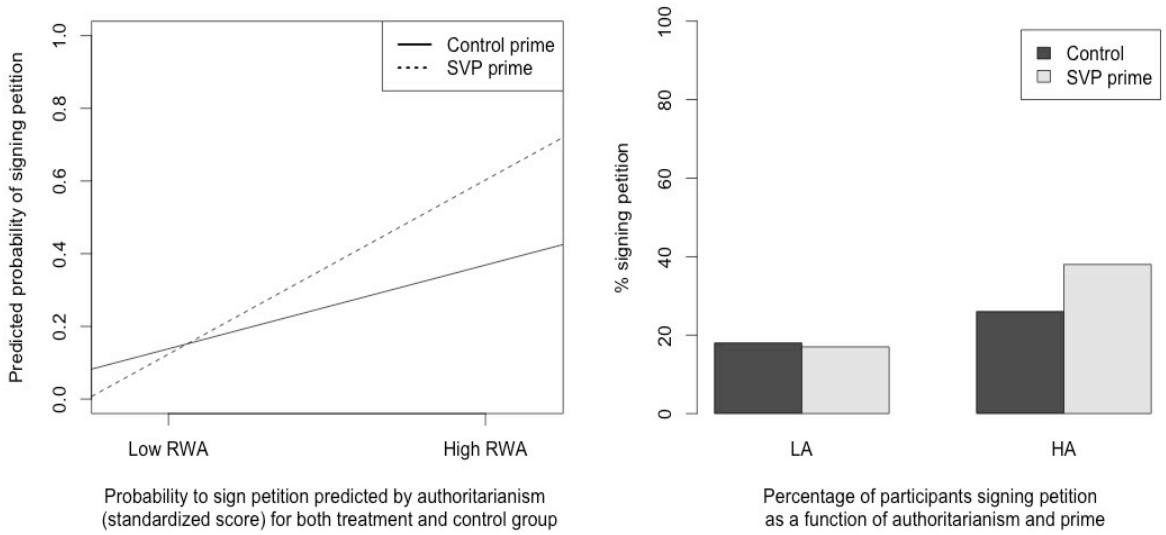

Figure 1.2: Results for Petition for Foreign Student Quota, Study 1

In sum, the results of Study 1 provide some, albeit limited, evidence that subliminal presentations of threatening political poster ads may influence the political behavior of high authoritarian individuals. Several aspects may, however, have reduced the chances of finding effects of the subliminal poster ad. First, if the effects of subliminal fear ads tend to be small – and the results point in this direction – , they simply require more statistical power and thus larger sample sizes to be detected. Second, it should be noted that the sample in Study 1 consisted of university students, a population known for being rather anti-authoritarian. Indeed, the sample mean of authoritarianism was at a low 32, and varied only little (SD = 9.9) compared to Study 2 (SD = 14.6). In other words, the student sample that was used in Study 1 did not provide enough high authoritarian observations and it was thus unlikely to produce the effects that I expected based on theoretical grounds. Third, it might be that movies are not an ideal container for subliminal priming or advertising. In fact, most of the studies in which subliminal advertising proved effective relied on priming tasks (e.g., lexical decision task) in order to subliminally prime subjects. In Study 2, I took these considerations into account by conducting an online experiment using Amazon's Mechanical Turk system. This procedure allowed for a replication of Study 1 in a more natural study environment, the recruitment and use of a larger sample, and the use of different subjects, primes, and prime containers.

1.5 Study 2

1.5.1 Participants and design

In exchange for $2.50, 478 (236 female, 242 male) non-Hispanic White Americans completed the study via Amazon's MTurk platform. This sample size was slightly higher than the one determined a priori on the basis of the desired power (.80), alpha level (.05), and anticipated effect size ($d = 0.26$; see Study 1).[6] Participants' mean age was 31.6 years. The study was a 2 (treatment vs. control) x 2 (low authoritarians vs. high authoritarians) between-subjects factorial design. The experimental conditions were published twice on MTurk at different time points (6 a.m. EDT/6 p.m. EDT) and in sequential random order. This procedure was chosen to reduce the chance of potential selection biases.

1.5.2 Procedure

After accepting the study (or "Human Intelligence Task" (HIT) in Amazon's lingo) on the MTurk web interface, subjects were forwarded to the Inquisit launch page. As on MTurk, they were told that the experiment included several questionnaires and a computer task. In the first part of the study, subjects were asked to provide socio-demographic data as well as information on their political orientation, party identification, and authoritarianism. In the second part, participants completed the subliminal priming task in which they were asked to discriminate between pleasant and unpleasant pictures (see Appendix 1.A) by pressing a left key ("E") for unpleasant and a right key ("I") for pleasant pictures. Prior to each target stimulus, the subliminal prime was flashed for approximately 17 ms. Each prime was preceded and followed by a mask for 50ms. The latter was created by decomposing the target stimuli and the treatment prime into 100 parts and randomly selecting from these elements to create the mask (see Figure 1.3). In order to mask the slogan of the prime (i.e., "No More Minarets"), red and black parts from the treatment stimulus were randomly selected and placed at the same position as in the treatment prime.

[6] Soper's (2013) sample size calculator indicated that the minimum required sample size for a two-tailed hypothesis test was 468.

Figure 1.3: Prime and Mask, Study 2

After a practice block of 10 trials, subjects completed another 40 trials for a total of 50 trials in which the subliminal stimulus was presented. In the third part, subjects' support of a ban on mosques and minarets was measured. Finally, in order to test whether the presentation of the primes was subliminal, a similar procedure was applied as in Study 1. That is, participants were informed that a picture had been primed and were asked to describe it. In addition, they were subjected to a forced-choice test in which they had to choose from a list of three primes (see Figure 1.4) the one that they thought had been primed. After the subliminality check, participants received a unique completion code to submit the HIT on the MTurk interface and were thanked for their participation.

Figure 1.4: Posters Used in Forced-Choice Test of Study 2

1.5.3 Independent variables

Prime condition
An Americanized version of the Swiss People's Party's (SVP) poster ad for a ban on minarets (see Figure 1.3) was primed for a duration of 17 milliseconds. The poster was used by the SVP in their political campaign against minarets in 2009. Both the campaign and the outcome of the vote (the initiative was accepted by a 57.5 percent majority) spurred a heated debate on the use and impact of fear campaigns in Swiss ballots.

Authoritarianism
Authoritarianism was measured using the same 12-item RWA-scale (Cronbach's $\alpha = 0.89$) as in Study 1. Again, the sample median (= 37) was used to split participants into low authoritarians (n = 247) and high authoritarians (n = 231).

1.5.4 Dependent variables

Support for a ban on mosques and minarets
Participants indicated their support for a ban on mosques and minarets on a single item ("Do you support or oppose a ban on mosques and minarets?" [1 = *strongly oppose*, 7 = *strongly support*]). Based on my theoretical model, I expected the subliminal prime to increase support, yet only among participants scoring high on right-wing authoritarianism.

1.5.5 Results and discussion

Subliminality check
None of the subjects reported seeing the poster prime. If anything, subjects could describe the composition and colors of the mask. In addition, subjects in the treatment condition (30%) were as likely ($t(476) = 0.18$, $p = .43$) to choose the correct prime as subjects in the control (31%) condition. Therefore, results did not indicate a violation of subliminal presentation.

Support for a ban on mosques and minarets
Descriptive statistics and the correlation between authoritarianism and the dependent variable are presented in Table 1.3. Consistent with my theoretical expectation, authoritarianism was significantly associated with support for a ban on mosques and minarets. In fact, support for the policy was much higher among high authoritarian individuals than among low authoritarian participants (see Table 1.4). A regression analysis for subjects' support of a ban on mosques and minarets on priming condition, authoritarianism, and their interaction revealed a main effect of the prime ($\beta = .41$, $t(477) = 3.41$, $p < .01$), a main effect of authoritarianism ($\beta = .55$, $t(477) = 6.42$, $p < .001$), and the expected interaction between priming condition and right-wing authoritarianism ($\beta = .31$, $t(477) = 2.65$, $p < .01$). I also ran an ordinal logistic regression analysis, which revealed both a main effect of the subliminal poster ad (OR: 1.68; CI: 1.10-2.55; $p < .05$) and authoritarianism (OR: 2.66; CI: 1.97-3.60; $p < .001$), but a

non-significant interaction (OR: 1.35; CI: 0.90-2.02; $p = .15$).

Table 1.3: Descriptive statistics and correlations between right-wing authoritarianism and the dependent variable (Study 2)

	M	**SD**	**RWA**	**Support for a ban on mosques and minarets**
RWA	38.99	14.62	-	
Support for a ban on mosques and minarets	1.95	1.54	0.49***	-

* p<.05, ** p<.01, *** p<.001.

Table 1.4: Support for a ban on mosques and minarets as a function of priming condition and right-wing authoritarianism (Study 2)

	Right-wing authoritarianism				**Priming condition**			
	Low (n=247)		**High (n=231)**		**Control (n=236)**		**Prime (n=242)**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Support for a ban on mosques and minarets	1.30	0.85	2.65	1.79	1.75	1.35	2.14	1.68

However, visualizations of the data (see Figure 1.5) clearly revealed the predicted interaction. The average support for a ban on mosques and minarets among low and high authoritarian participants in the two priming conditions is depicted in Figure 1.5 (right panel). Consistent with my expectation, high authoritarian participants in the poster prime condition exhibited significantly ($t(229) = -2.95$, $p < .01$) higher support for a ban on mosques and minarets ($M = 2.99$, $SD = 1.91$) than their counterparts in the neutral control condition ($M = 2.31$, $SD = 1.60$). Importantly, this effect was larger ($d = 0.39$) than the main effect of the prime ($d = 0.26$). Finally, the average support for a ban on mosques and minarets was only marginally ($t(245) = -1.61$, $p = .05$) higher among low authoritarians in the treatment condition ($M = 1.38$, $SD = 0.95$) than among those in the control condition ($M = 1.21$, $SD = 0.71$).

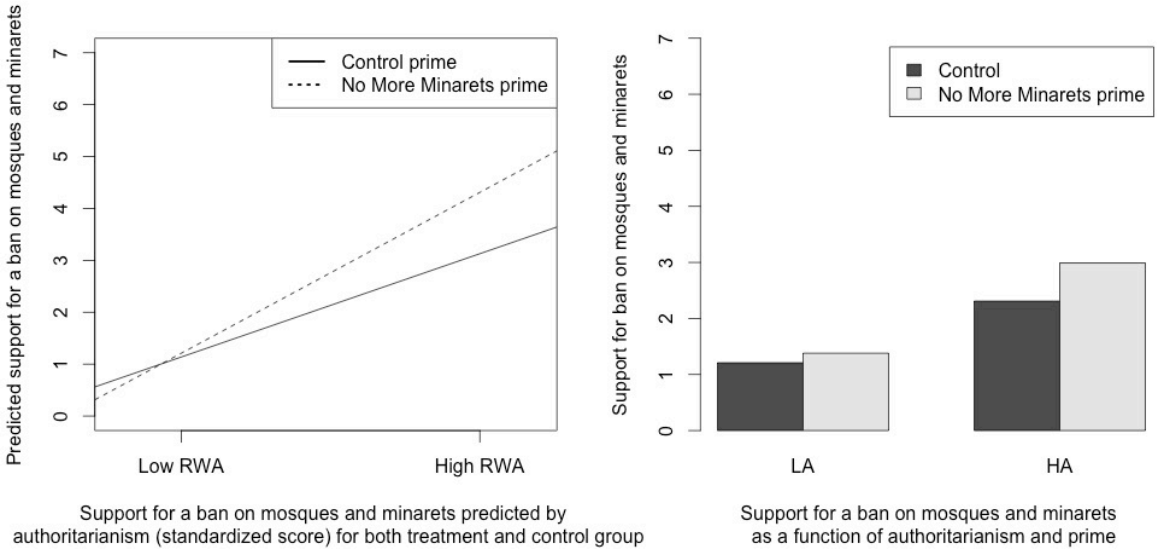

Figure 1.5: Results for Ban on Mosques and Minarets, Study 2

Taken together, findings of Study 2 corroborate my hypothesis that high authoritarian people are particularly susceptible to persuasion brought about by subliminal fear ads. As in the first study, authoritarianism predicted support for political action against a social outgroup and the effect was moderated by exposure to the subliminal poster prime. However, in contrast to the previous study, results from the second study also suggest a main effect of the prime. That is, even low authoritarian individuals may be affected by subliminal versions of right-populist fear ads, albeit to a weaker degree than high authoritarian subjects.

1.6 General Discussion

In the present article, I have argued that the necessary conditions of subliminal advertising are satisfied in the case of high authoritarian individuals if the latter are exposed to the type of threat appeals typically used in right-populist fear campaigns. I have argued in this manner because the policies that these campaigns promote are consistent with the political goals of high authoritarian individuals (i.e., campaign posters are goal-relevant cognitions). Evidence from two studies in which participants were exposed to subliminal political poster ads supports my hypothesis. In Study 1, a subliminal poster against immigration increased the probability that participants would sign a petition for a foreign student quota, yet only among high authoritarian participants. The same pattern emerged in Study 2 where a poster against minarets did have a stronger effect on high authoritarian subjects' support for a ban on mosques and minarets than on the support among low authoritarians. Within the high

authoritarian condition, the found effects varied between $d = 0.26$ (Study 1) and $d = 0.39$ (Study 2). As such, they were smaller than the effects found in Karremans et al. (2006) (e.g., Study 2: d=0.76) and Bermeitinger et al. (2009) (d=0.80)[7] but larger than the average effect found in Trappey's (1996) meta-analysis (d=0.11). What accounts for this intermediate position? First, upon examination of the different domains in which the studies were run, it is very likely that political goals are not as fundamentally important as physical (e.g., drinking) or mental needs (e.g., concentration). It would, in fact, be naive to consider the goal of keeping society free of immigrants as an equally urgent goal as needing to quench one's thirst. After all, thirsty authoritarians would still start by quenching their thirst, killing their fatigue second, and voting for a ban on immigration third. Therefore, even if the conditions of subliminal advertising are met, the likelihood and strength of subliminal persuasion will still vary strongly across different domains. Second, while the research reviewed in Trappey (1996) was carried out decades ago, the present studies have benefited from very recent insights into the workings of subliminal advertising. As such, they were better able to address the necessary conditions of subliminal advertising and thereby create a framework in which subliminal effects could indeed occur.

My results have important implications. First, although effects were generally weak, they are nevertheless existent and worth our attention. Recall that both poster ads were in fact used in Swiss politics and ultimately allowed the Swiss people's party to win their campaigns against immigration and minarets. Although broadcast political advertising is illegal in Switzerland, attempts to subliminally persuade voters may still be of interest to political parties and organizations. This may, in particular, be true for right-populist party events where high authoritarian voters are likely to show up in large numbers. In such a place, people's motivational goals will be in line with subliminal poster ads promoting anti-immigrant or anti-Islam policies and, at least some of them, may thus be won over by subliminal priming. Another potential channel for subliminal advertising in politics are social media platforms, which may still be abused for subliminal messaging. Second, the results from Study 1 suggest that not only political attitudes may be influenced by subliminal primes but also real political behavior. After all, subjects in Study 1 signed a petition that was actually sent to the head of the university. Thus, if political actors can combine subliminal exposure to fear ads with the collection of signatures, they may well reduce the time it takes them to achieve the number of signatures required for submission of the referendum or petition. The same, of course, applies to situations where subliminal advertising can be administered in close proximity to voting booths. It is left for future research to investigate the temporal persistence of subliminal persuasion in the domain of political fear campaigns. Also, future research should examine whether the subliminal advertising paradigm also applies to other political domains and concepts. If it does, not only right-populist parties but also other parties may have a tool to secretly persuade their followers.

[7]The reported d statistic reflects the average of all 2 (subliminal prime: logo A vs. logo B) x 2 (order of bowl arrangement: AB vs. BA) = 4 comparisons among tired participants (Bermeitinger et al. 2009, 324)

References

Adorno, T.W., Frenkel-Brunswick, E., Levinson, D.J., and Nevitt Sanford. 1950. The authoritarian personality. New York: Harper and Row.

Altemeyer, Bob. 1996. The Authoritarian Specter. Cambridge: Harvard University Press.

Arendt, Florian. 2013. Dose-dependent media priming effects of stereotypic newspaper articles on implicit and explicit stereotypes. *Journal of Communication*, 63, 830-851.

Arendt, F., Marquart, F., and Jörg Matthes. 2015. Effects of right-wing populist political advertising on implicit and explicit stereotypes. *Journal of Media Psychology*. Posted online. doi: 10.1027/1864-1105/a000139

Bargh, J.A., and Tanya L. Chartrand. 2000. The mind in the middle: A practical guide to priming and automaticity research. In Reis H., and C. Judd (Eds.), *Handbook of research methods in social and personality psychology*, 253-285. New York: Cambridge University Press.

Bermeitinger, C., Goelz, R., Johr, N., Neumann, M., Ecker, U.K.H., and Robert Doerr. 2009. The hidden persuaders break into the tired brain *Journal of Experimental Social Psychology*, 45(2), 320-326.

Betz, Hans-Georg. 2013. Mosques, minarets, burqas, and other essential threats: The populist right's campaign against Islam in Western Europe. In Wodak, Ruth, KhosraviNik, Majid, and Brigitte Mral (Eds.), *Right-wing populism in Europe: Politics and discourse*, 71-88. London: Bloomsbury.

Brader, Ted. 2006. Campaigning for hearts and minds: How emotional appeals in political ads work. Chicago: University of Chicago Press.

Butler, J. Corey 2000. Personality and Emotional Correlates of Right-Wing Authoritarianism. *Social Behavior and Personality*, 28(1), 1-14.

Byrne, Donn. 1959. The effect of a subliminal food stimulus on verbal response. *Journal of Applied Psychology*, 43, 249-252.

Caccavale, J.G., Wanty, T.C., and Julie A. Edell. 1981. Subliminal implants in advertisements: An experiment. *Advances in Consumer Research*, 9, 418-423.

Champion, J.M., and Weld W. Turner. 1959. An experimental investigation of subliminal perception. *Journal of Applied Psychology*, 43(6), 382-384.

Cohen, Jacob. 1988. Statistical power analysis for the behavioral sciences (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cohrs, J.C., Moschner, B., Maes, J., and Sven Kielmann. 2005. The Motivational Bases of Right-Wing Authoritarianism and Social Dominance Orientation: Relations to Values and Attitudes in the Aftermath of September 11, 2001. *Personality and Social Psychology Bulletin*, 31(10), 1425-1434.

Cooper, J., and Grant Cooper. 2002. Subliminal Motivation: A Story Revisited. *Journal of Applied Social Psychology*, 32, 2213-2227.

Cuperfain, R., and Keith T. Clarke. 1985. A new perspective of subliminal perception. *Journal of Advertising*, 14, 36-41.

Doty, R.M., Peterson, B.E., and David G. Winter. 1991. Threat and authoritarianism in the United States, 1978-1987. *Journal of Personality and Social Psychology*, 61, 629-640.

Duckitt, John. 1989. Authoritarianism and Group Identification: A new View of an Old Construct. *Political Psychology*, 10(1), 63-84.

Duckitt, John. 2006. Differential Effects of Right-Wing Authoritarianism and Social Dominance Orientation on Outgroup Attitudes and Their Mediation by Threat From and Competitiveness to Outgroups. *Personality and Social Psychology Bulletin*, 32(5), 684-696.

Feldman, S., and Karen Stenner. 1997. Perceived Threat and Authoritarianism. *Political Psychology*, 18(4), 741-770.

Feldman, S., Lavine, H., Lodge, M., and Bradley Verhulst. 2010. Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli. Paper presented at the annual meeting of the International Society of Political Psychology, San Francisco, CA.

Fromm, Erich. 1941. Escape from Freedom. New York: Holt, Rinehart and Winston.

Funke, Friedrich. 2005. The Dimensionality of Right-Wing Authoritarianism: Lessons from the Dilemma between Theory and Measurement. *Political Psychology*, 26(2), 195-218.

Gable, M., Wilkens, H.T., Harris, L., and Richard Feinberg. 1987. An Evaluation of Subliminally Embedded Sexual Stimuli in Graphics. *Journal of Advertising*, 16(1), 26-31.

Gawronski, B., and Galen V. Bodenhausen. 2006. Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692-731.

George, S.G., and Luther B. Jennings. 1975. Effect of Subliminal Stimuli on Consumer Behavior: Negative Evidence. *Perceptual and Motor Skills*, 41, 847-854.

Hassin, R.R., Ferguson, M.J., Shidlovski, D., and Tamar Gross. 2007. Subliminal exposure to national flags affects political thought and behavior. Proceedings of the National Academy of Sciences PNAS, 104(50), 19757-19761.

Hassin, R.R., Ferguson, M.J., Kardosh, R., Porter, S.C., Carter, T.J., and Veronika Dudareva. 2009. Precis of Implicit Nationalism. *Annals of the New York Academy of Sciences*, 1167, 135-145.

Hawkins, Del. 1970. The Effects of Subliminal Stimulation on Drive Level and Brand Preference. *Journal of Marketing Research*, 7(3), 322-326.

Karremans, J.C., Stroebe, W., and Jasper Claus. 2006. Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. *Journal of Experimental Social Psychology*, 42, 792-798.

Kilbourne, W.E., Painton, S., and Danny Ridley. 1985. The Effect of Sexual Embedding on Responses to Magazine Advertisements. *Journal of Advertising*, 14(2), 48-56.

Lodge, M., and Chuck Taber. 2005. The automaticity of affect for political leaders, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, 26(3), 455-482.

Manganelli Rattazzi, A.M., Bobbio, A., and Luigina Canova. 2007. A short version of the Right-Wing Authoritarianism (RWA) Scale. *Personality and Individual Differences*, 43, 1223-1234.

Peterson, B.E., Doty, R.M., and David G. Winter. 1993. Authoritarianism and Attitudes Toward Contemporary Social Issues. *Personality and Social Psychology Bulletin*, 19(2), 174-184.

Pratkanis, Anthony. 1992. The cargo-cult science of subliminal persuasion. *Skeptical Inquirer*, 16, 260-272.

Rokeach, Milton. 1960. The open and closed mind. New York: Basic Books.

Sales, Stephen M. 1972. Economic threat as a determinant of conversion rates in authoritarian and nonauthoritarian churches. *Journal of Personality and Social Psychology*, 23(3), 420-428.

Soper, Daniel S. 2013. A-priori Sample Size Calculator for Student t-Tests. http://www.danielsoper.com/statcalc ([Software]. Available from)

Stenner, Karen. 2005. The Authoritarian Dynamic. Cambridge: Cambridge University Press.

Stewart, P.A., and James N. Schubert. 2006. Taking the "Low Road" with Subliminal Advertisements: A Study Testing the Effect of Precognitive Prime "RATS" in a 2000 Presidential Advertisement. *The Harvard International Journal of Press/Politics*, 11(4), 103-114.

Strahan, E.J., Spencer, S.J., and Mark P. Zanna. 2002. Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology*, 38, 556-568.

Thomsen, L., Green, E.G.T., and Jim Sidanius. 2008. We will hunt them down: How social dominance orientation and right-wing authoritarianism fuel ethnic persecution of immigrants in fundamentally different ways. *Journal of Experimental Social Psychology*, 44, 1455-1464.

Trappey, Charles. 1996. A meta-analysis of consumer choice and subliminal advertising. *Psychology and Marketing*, 13(5), 517-530.

Verwijmeren, T., Karremans, J.C., Stroebe, W., and Daniel H.J. Wigboldus. 2011. The workings and limits of subliminal advertising: The role of habits. *Journal of Consumer Psychology*, 21, 206-213.

Wegener, D.T., and Richard E. Petty. 1997. The flexible correction model: The role of na ve theories of bias in bias correction. In Zanna, Mark (Ed.), *Advances in experimental social psychology*, 141-208. New York: Academic Press.

Weinberger, J., and Drew Westen. 2008. RATS, We Should Have Used Clinton: Subliminal Priming in Political Campaigns. *Political Psychology*, 29(5), 631-651.

Weinstein, S., Drozdenko, R., and Curt Weinstein. 1986. Effects of subliminal cues in print advertising upon brain response, purchase intention, and simulated purchase. In Olson, Jerry C., and Keith Sentis (Eds.), *Advertising and Consumer Psychology*, proceedings from the Advertising and Consumer Psychology Conference 3, 3-16.

Wilson, Glenn D. 1973. The Psychology of Conservatism. New York: Academic Press.

Appendix

1.A Stimuli List for Subliminal Priming Task

Figure 1.A.1: Pleasant pictures used in subliminal priming task (Study 2)

Figure 1.A.2: Unpleasant pictures used in subliminal priming task (Study 2)

Perhaps the most compelling way for implicit social cognition to establish its relevance to the study of politics is to enhance researchers' ability to predict political behavior.

(Nosek et al. 2010, 558)

2

Single-Target Implicit Association Tests (ST-IAT) Predict Voting Behavior of Decided and Undecided Voters in Swiss Referendums

Livio Raccuia[1]

UNDECIDED voters represent a major challenge to political pollsters. Recently, political psychologists have proposed the use of implicit association tests (IAT) to measure implicit attitudes toward political parties and candidates and predict voting behavior of undecided voters. A number of studies have shown that both implicit and explicit (i.e., self-reported) attitudes contribute to the prediction of voting behavior. More importantly, recent research suggests that implicit attitudes may be more useful for predicting the vote of undecided voters in the case of specific political issues rather than elections. Due to its direct-democratic political system, Switzerland represents an ideal place to investigate the predictive validity of IATs in the context of political votes. In this article, I present evidence from three studies in which both explicit and implicit attitudes were measured ahead of the vote on four different referendums. Explicit attitudes predicted voting better than implicit attitudes for decided voters while implicit and explicit attitudes were equally good predictors among undecided voters. In addition, implicit attitudes predicted voting behavior descriptively, but not significantly better for undecided voters while, also from a descriptive point of view, explicit attitudes predicted voting better for decided respondents. In sum, results suggest that, as argued in previous research, the predictive value of implicit attitudes may be higher in the context of issue-related votes but still not as high as initially hoped-for.

[1]This chapter is based on the following article: Raccuia, L. (2016). Single-Target Implicit Association Tests (ST-IAT) Predict Voting Behavior of Decided and Undecided Voters in Swiss Referendums. PLoS ONE 11(10): e0163872. doi: 10.1371/journal.pone.0163872

2.1 Introduction

Explaining the political behavior of people lies at the heart of political research. In fact, opinion and exit polls have become an indispensable part of contemporary democracies. However, when asked to indicate their voting intention, people may either reveal which party or candidate they plan to vote for, or they may report themselves as being undecided. Irrespective of whether they are in fact undecided or simply unwilling to report their voting intention, undecided voters represent a major source of uncertainty when it comes to predicting the outcome of an election or vote. For this reason, political scientists have been looking for ways to improve polling accuracy by allocating undecided voters to the respective candidates, parties, or political camps. While many of these attempts have been based on using voter registration data (Visser et al. 2000), more recent attempts have focused on using implicit attitudes in order to improve the measurement and prediction of political behavior. Unlike explicit (i.e., self-reported) attitudes, implicit attitudes are more likely to operate below conscious awareness. In fact, previous research (e.g., Gawronski et al. 2006) has demonstrated that although people may be aware of their implicit attitudes, they may not be aware of (and thus not necessarily endorse) the ways in which their implicit preferences may impact their behavior. Hence, in the political realm, individuals' implicit attitudes toward parties or candidates may affect their eventual voting behavior beyond their explicit attitudes. To date, the most widely used method to measure implicit attitudes is the implicit association test (IAT) (Greenwald et al. 1998). While IATs have been used in many different domains, they were found to perform particularly well for the prediction of political behavior (Greenwald et al. 2009). In fact, IATs were shown to provide predictive validity over and above explicit measures in a number of studies conducted ahead of political elections (Arcuri et al. 2008; Friese et al. 2007; Roccato and Zogmaister 2010). Therefore, evidence strongly suggests that political pollsters should take into account both explicit and implicit measures when concerned with the prediction of voting behavior.

In this article, I build on these and more recent findings (e.g., Friese et al. 2012; Galdi et al. 2008) suggesting that IATs may be more useful for the prediction of voting behavior in the case of specific political issues (rather than political elections) as a result of less elaborated attitudes toward these issues. If this is true, then IATs should be useful tools for the prediction of political outcomes in Switzerland, where people frequently decide on issue-related referendums. In contrast to political elections where voting behavior is typically determined by long-lasting party affiliations, voting on referendums represents a much more complex situation. In Swiss federal votes, coalitions often span across traditional party lines and voters frequently deal with highly complex issues. As a result, attitudes toward these issues most likely tend to be less elaborated than attitudes toward political parties and candidates. Somewhat surprisingly, however, no efforts have yet been made to investigate the predictive validity of IATs for vote outcomes in Swiss referendums. For this reason, I have first assessed the predictive validity of different types (traditional computer vs. computer-

administered paper-format) of single-target implicit association tests (ST-IAT) in two studies that were conducted ahead of votes on such diverse issues as the purchase of fighter jets, minimum wage implementation, and public health insurance. In a second step, I have investigated whether implicit attitudes can add to the explanation of voting behavior of decided and undecided voters in the case of a very controversial referendum on immigration. All three studies were conducted online. In what follows, I will outline the most recent findings on the use of implicit attitudes for the prediction of political behavior and then present evidence from all three studies.

2.2 Implicit Attitude Measures and the Prediction of Political Behavior

At first, implicit measures were seen as fit instruments to predict spontaneous, uncontrolled behavior, while rather unfit for predicting behavior in situations where people engage in deliberation and thus exert control over their behavior (e.g., Dovidio et al. 1997). As such, the value of IATs was deemed rather low for the prediction of political behavior. However, the focus soon shifted to mixture or additive models in which both explicit and implicit measures were believed to uniquely contribute to the explanation of (political) behavior (e.g., Perugini 2005). In the political domain, Karpinski et al. (2005) were first to use IAT scores to predict voting intention in the 2000 U.S. presidential election. They found IAT scores to be a significant predictor of voting intention. However, implicit attitudes no longer predicted voting intention once they controlled for explicit attitudes such as self-reported liking of the two candidates. Concerning the prediction of actual voting behavior, Friese et al. (2007) used IATs for the first time in the lead-up to the 2002 parliamentary elections in Germany and found incremental predictive validity of Single-Target IATs (ST-IAT) for the five major political parties. Although self-reported party preferences were stronger predictors of voting intention and actual voting behavior, the implicit measures added significantly to the explanation of voting behavior. Hence, their findings were the first evidence in favor of additive prediction models in the domain of politics. In yet another study conducted in the lead-up to national elections, Arcuri and colleagues (2008) also found predictive validity of IATs. More importantly, they found IATs to be significant predictors of voting behavior among both decided and undecided voters. Their findings were highly intriguing, as they increased optimism in the use of implicit measures for voters who do not − for whatever reason − report their voting intention. In a parallel study, Galdi et al. (2008) first tested the utility of IATs in predicting future opinion and choice in the case of a local political issue in Italy. They found that for participants who were initially undecided, future (i.e., one week later) opinion about the enlargement of a U.S. military base was significantly predicted by a measure obtained in a Single-Category IAT (SC-IAT) (Karpinski and Steinman 2006). Explicit attitudes, however, did not predict future opinions. Interestingly, for decided subjects, the pattern was exactly reversed. Against the background of their results, Galdi and colleagues (2008) discussed convincing explanations for why implicit attitudes may be particularly useful for

predicting the behavior of undecided individuals: while in the process of decision-making, the implicit attitude leads individuals to selectively expose themselves to information that corresponds with their implicit attitudes. As a result of the biasing influence of implicit attitudes, undecided voters may eventually develop a conscious (i.e., explicit) preference for one political party or candidate and this may then determine their ultimate vote decision. In addition, implicit attitudes may also predict voting behavior independent of their role in forming explicit attitudes. For example, implicit attitudes may lead to biased interpretations of information such as voting options even at the time of the vote. Interestingly, the authors successfully demonstrated their biased-processing account in a follow-up study (Galdi et al. 2012).

In two recent studies, Roccato and Zogmaister (2010) and Friese et al. (2012) used IATs in the context of either national or presidential elections in Italy, Germany, and the United States. While both studies report evidence in favor of the additive prediction model, Friese and colleagues (2012) found that explicit attitudes were better predictors of actual vote choice for both decided and undecided voters, while implicit attitudes were better predictors for the voting behavior of decided rather than undecided voters. As these findings are in contrast to Galdi et al.'s (2008) results, the authors seek an explanation by pointing to the moderating role of cognitive elaboration. They argue that the more elaborated attitudes are, the less potential exists for implicit attitudes to explain variance over and above explicit attitudes. In fact, while elaboration may be lower in the case of specific political issues as in the Galdi et al. (2008) study, it is rather high in the case of political party or candidate preferences. To further corroborate their claim, they show that correlations between their implicit and explicit measures were much stronger than in the case of Galdi and colleagues' study (2008), another indicator of high cognitive elaboration (see also Karpinski et al. 2005; Roccato and Zogmaister 2010). Finally, Lundberg and Payne (2014) used the Affect Misattribution Procedure (AMP) to capture implicit attitudes and a more precise measure of decidedness than the typically used dichotomous operationalization. Concretely, they used a continuous measure for confidence in one's voting intention (1="not sure at all" to 5="extremely sure"), finding that at high levels of confidence, explicit attitudes are much stronger predictors than implicit attitudes, while at low levels of confidence, implicit and explicit attitudes are equally predictive of voting behavior. Moreover, Lundberg and Payne (2014) provide a strong argument for why implicit attitudes may help to foresee the behavior of undecided voters: they show that confidence in one's vote is a function of the strength of one's explicit attitude toward the party or candidate. Hence, when individuals report themselves as either decided or undecided, they draw upon their explicit attitudes, neglecting their implicit attitudes. This, in turn, explains why explicit measures were found to be better predictors for the voting behavior of decided voters (Friese et al. 2012; Galdi et al. 2008; Lundberg and Payne 2014), while implicit measures were found to be good predictors for the voting behavior of both decided and undecided voters (Arcuri et al. 2008; Lundberg and Payne 2014).

Taken together, there is converging evidence (e.g., Friese et al. 2007, 2012; Lundberg and Payne 2014; Roccato and Zogmaister 2010) in favor of additive prediction models that take

into account both explicit and implicit measures. In other words, both explicit and implicit measures should be considered when predicting the outcome of political elections or votes. However, there is also convincing evidence on when implicit attitudes may be particularly useful and when they are not. In several studies (e.g., Friese et al. 2012; Karpinski et al. 2005; Lundberg and Payne 2014; Roccato and Zogmaister 2010) conducted before political elections, the incremental predictive validity of IATs was relatively small as a result of rather high correlations between explicit and implicit measures. As discussed in Friese et al. (2012), high consistency between the two types of measures can be explained with high cognitive elaboration. Because party or candidate preferences are well-elaborated attitudes, (ST-)IATs may be less attractive for predictions in the case of political elections. Yet, they may still be useful to predict voting behavior, particularly among undecided voters, in the case of specific political issues (see Galdi et al. 2008). Due to its direct democratic political system, Switzerland provides a perfect setting for testing the predictive validity of (ST-)IATs in the context of national referendums. To my knowledge, however, no efforts have been made as yet in this direction. For all these reasons, I first examined two different types of ST-IATs and then selected one of them to assess its predictive validity for both decided and undecided voters in the case of a crucial referendum in Switzerland.

2.3 Study 1

The first study was conducted between April 28, 2014 and May 7, 2014 and included two ST-IATs for assessing respondents' implicit attitudes toward two referendums: first, a popular initiative for implementing a national minimum wage, and second, a referendum about the purchase of 22 Gripen fighter jets worth 3.1 billion Swiss francs. The vote was held on May 18, 2014. Both referendums were rejected (minimum wage initiative: 76.3 percent; Gripen referendum: 53.4 percent). The main goal of Study 1 was to get a first hint on whether ST-IATs can be used for the prediction of voting behavior in Swiss referendums. To this end, a small student sample was deemed sufficient and no power analysis was conducted to determine sample size. Instead, the survey was kept active until no additional responses were registered. Furthermore, because a pretest had shown that correspondence between voting intention and voting behavior is high ($r = .81$), a one-wave design was chosen. That is, subjects were only surveyed before the vote and asked to report their voting intention or vote choice if they had already voted.

2.3.1 Ethics statement

The study was conducted online. All subjects were older than 18 years of age. In the invitation, subjects received brief information about the topic (Swiss Federal Votes) and the duration of the study as well as the researcher involved and his affiliation. On the first page, subjects were informed about the measures involved (questionnaire and two implicit association tests). At the end of the study, subjects received contact information. All data was

analyzed anonymously and no identifying data (e.g., names) were collected. At the author's faculty (Faculty of Arts, University of Zurich) the IRB asks researchers to self-assess the ethical soundness of their research using a checklist and only seek approval from the IRB if one of the questions is answered affirmatively. This was not the case for the present study.

2.3.2 Participants

Participants were 268 (47.4 % males, 52.6 % females) students, doctoral students, and employees of the University of Zurich. They received an email invite and were told that the study was about the upcoming national votes. Furthermore, they were asked to participate only if they are entitled to vote. Subjects' mean age was 30.90 years (SD=10.03) and they received no compensation in exchange for their participation.

2.3.3 Procedure

The entire study was written and administered using the Inquisit software and hosted at http://www.millisecond.com. After completing a short questionnaire on socio-demographic questions and questions pertaining to political interest and partisanship, participants completed the two ST-IATs in randomized order.

2.3.4 Measures

Implicit attitudes toward the two referendums were measured using a similar ST-IAT as the one used in Friese et al. (2007) (see also Bluemke and Friese 2008; Dotsch et al. 2011). However, unlike their ST-IAT, I used an additional block of 15 trials ahead of each critical block, leading to a total number of 50 trials for each combination (see Table 2.1). This procedure was chosen because of the results from a pretest in which the predictive validity of the ST-IAT was unsatisfactory due to an order effect that occurred as a result of the low number of trials. As in Friese et al. (2007), however, block and stimulus order were kept constant across respondents with each respondent starting with the positive vs. negative AND referendum response set. As can be seen in Table 1, subjects started with a practice block of 20 trials in which they had to discriminate between positive and negative words. Next, they completed two blocks of 15 and 35 trials in which positive attributes were assigned to the left-key (i.e., "E") and negative attributes and words representing the respective referendum were assigned to the right-key (i.e., "I"). Finally, in the fourth and fifth block, the combination was reversed such that positive words and political stimuli were assigned to the left-key, while the right-key was left for negative words. Both evaluative categories were represented by five words (e.g., joy, love, stink, poison) while the two referendums were represented by three pictures (e.g., campaign poster) and two text stimuli (e.g., Swiss army) (see Appendix 2.A). For calculating the ST-IAT scores, I used the improved scoring algorithm D2 proposed by Greenwald et al. (2003). Thus, all trials with response latencies less than 400 ms or greater than 10 seconds were discarded from the analyses, and a built-in error penalty was used in

the case of incorrect responses. The resulting score can take on values between -2 and +2 and has several advantages. First, it provides a measure for both the direction and strength of one's implicit attitude. Second, it applies a penalty for respondent errors and thus for very low latencies produced when sorting errors are committed. Finally, by dividing the difference in block means by its corresponding pooled standard deviation, the algorithm takes the respondent's variability of response latencies into account (Greenwald et al. 2003). The order of the two ST-IATs was randomized across participants.

Table 2.1: Sequence of Trial Blocks in ST-IATs in Study 1

Block	No.of Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Attribute discrimination	Positive words	Negative words
2	15	Practice Combined Block 1	Positive words	Negative words / Gripen referendum / Minimum wage initiative
3	35	Combined Block 1	Positive words	Negative words / Gripen referendum / Minimum wage initiative
4	15	Practice Combined Block 2	Positive words / Gripen referendum / Minimum wage initiative	Negative words
5	35	Combined Block 2	Positive words / Gripen referndum / Minimum wage initiative	Negative words

Notes: The ST-IAT score is based on data from blocks 2,3,4 and 5. Block order was fixed across participants. For stimuli presentation, a fixed random order was used.

Explicit attitudes were assessed using an 11-point scale for political orientation (0 = *extreme left* to 10 = *extreme right*) with higher values indicating a more rightist placement. It turned out that both voting for the minimum wage initiative (r = -.66, p <.001) and the Gripen referendum (r = .48, p <.001) were closely related to participants' political orientation.

Voting intention was measured using the following question: "What do you plan to vote at (i) the popular initiative 'for a national minimum wage' (ii) the 'Gripen referendum'?" Response options were 'Yes', 'No', 'Don't know'. Undecided participants ($N_{Min.wage}$ = 26, N_{Gripen} = 17) were excluded from the analyses. This left 154 decided voters for the minimum wage initiative and 163 decided voters for the Gripen referendum. Intentions to vote against the referendum were coded as 0, and intentions to vote in favor of it were coded as 1.

Voting behavior was measured with the following question: 'How did you vote at (i) the popular initiative 'for a national minimum wage' (ii) the 'Gripen referendum'?" Response options were 'Yes', 'No', 'Abstained'. For both referendums, 86 participants reported their vote choice (there were no abstentions). Votes against the referendum were coded as 0, and votes in favor of it were coded as 1.

2.3.5 Results and discussion

Preliminary analyses

For each ST-IAT, I calculated the Spearman-Brown corrected split-half reliability as the correlation between the ST-IAT score of blocks 2 and 4 with the ST-IAT score of blocks 3 and 5. This led to similar reliability estimates for the minimum wage initiative (r = .57) and the Gripen referendum (r = .52). In addition, I examined the correspondence between the implicit and explicit measure for both referendums. In previous research (e.g., Friese et al. 2012; Karpinski et al. 2005), high implicit-explicit correspondence has been explained with well-elaborated attitudes. As a consequence, strong correlations between implicit and explicit attitudes were associated with high predictive validity of implicit attitudes overall but only little predictive value over and above explicit measures. Interestingly, I found only moderate correlations between the implicit and explicit measure for both the minimum wage initiative (r = -.40) and the Gripen referendum (r = .36). Correlations were stronger among participants who had already voted on the minimum wage initiative (r = -.48) than among those who had not yet voted but at least developed a voting intention (r = -.37). In the case of the Gripen referendum, these correlations were very similar across both groups of voters (Already voted: r = .33, Decided: r = .38). Taken together, these findings are consistent with the argument made in previous research (e.g., Friese et al. 2012, 2016) that attitudes toward political issues are less elaborated than attitudes toward political parties or candidates.

Voting intention and voting behavior

Due to the one-wave design of the study, there is no data on undecided voters' eventual voting decision. For this reason, I could not run a separate analysis on undecided voters. Therefore, in this very preliminary study, I cannot speak to the question of whether implicit attitudes are better predictors for undecided than decided voters. However, I can still assess the predictive value of implicit attitudes for the prediction of decided respondents' voting intention and behavior in Swiss federal votes in general and beyond explicit measures.

Table 2.2 shows that for both referendums, implicit attitudes (models 1a) predicted voting intention. In fact, implicit attitudes allowed for the correct classification (%CCC) of 85.1 percent of (decided) voters in the case of the Gripen referendum, but they only predicted 62.8 percent of cases correctly in the case of the minimum wage initiative. Explicit attitudes (models 1b) outperformed implicit attitudes on both occassions, correctly classifying 81.8 percent of cases (minimum wage initiative) and 85.3 percent of cases (Gripen referendum). Thus, it seems that consistent with previous research (e.g., Friese et al. 2012), explicit attitudes are better predictors for decided voters than implicit attitudes. The picture is, however, less clear if one looks at the incremental validity of the implicit measure. Interestingly, implicit attitudes did not predict voting intention for the minimum wage initiative beyond the explicit measure. They decreased Nagelkerke's R^2 by 0.5 percentage points after controlling for explicit attitudes (model 2) and they did not change the percentage of correctly classified cases. However, in the case of the Gripen referendum, implicit attitudes increased Nagelkerke's R^2

by 9.2 percentage points after accounting for explicit attitudes, while the %CCC increased by 3.9 percentage points.

Table 2.2: Results of logistic regression for prediction of voting intention (Study 1)

Step	**Variable**	**B**	***SE***	**Wald**	***p***	**Exp(B)**	**Nagelkerke's R^2**	**% CCC**
Minimum wage initiative (N=154)								
1a	Constant	.263	.185	1.425	.154	1.301	.167	62.8 %
	ST-IAT	.780	.200	3.902	<.001	2.181		
1b	Constant	.117	.215	.546	.585	1.124	.511	81.8 %
	Explicit measure	-2.057	.359	-5.730	<.001	.128		
2	Constant	.197	.227	.868	.385	1.218	.506	81.8 %
	ST-IAT	.306	.241	1.269	.204	1.358		
	Explicit measure	-1.781	.367	-4.849	<.001	.168		
Gripen referendum (N=163)								
1a	Constant	-1.885	.295	-6.393	<.001	.152	.321	85.1 %
	ST-IAT	1.442	.300	4.802	<.001	4.229		
1b	Constant	-1.727	.260	-6.638	<.001	.178	.398	85.3 %
	Explicit measure	1.419	.245	5.798	<.001	4.131		
2	Constant	-1.990	.319	-6.241	<.001	.137	.490	89.2 %
	ST-IAT	1.015	.324	3.130	.002	2.760		
	Explicit measure	1.168	.285	4.098	<.001	3.216		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting intention (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

Table 2.3 depicts the results for the prediction of voting behavior. As for voting intention, explicit attitudes better predicted vote choice in the case of the minimum wage initiative. They allowed for the correct classification of 81.4 percent of cases compared to only 69.4 percent of cases that were correctly predicted by implicit attitudes. However, for the Gripen referendum, the implicit measure did a slightly better job than explicit attitudes in terms of correctly classified cases. It correctly predicted 82.4 percent of voters while the explicit measure did so for 81.4 percent of voters. Finally, for the latter referendum, adding the implicit to the explicit measure led to an increase in Nagelkerke's R^2 by 8.7 percentage points and an increase in the %CCC of 1.0 percentage points. In contrast, implicit attitudes did not predict voting on the minimum wage initiative beyond the explicit measure (-2.2 %CCC).

Table 2.3: Results of logistic regression for prediction of voting behavior (Study 1)

Step	Variable	B	*SE*	Wald	*p*	Exp(B)	Nagelkerke's R^2	% CCC
Minimum wage initiative (N=86)								
1a	Constant	.356	.265	1.345	.179	1.400	.243	69.4 %
	ST-IAT	1.051	.323	3.256	.001	2.860		
1b	Constant	.262	.298	.879	.379	1.300	.555	81.4 %
	Explicit measure	-2.167	.468	-4.630	<.001	.115		
2	Constant	.073	.363	.200	.841	1.075	.615	79.2 %
	ST-IAT	.405	.407	.994	.320	1.499		
	Explicit measure	-2.487	.697	-3.568	<.001	.083		
Gripen referendum (N=86)								
1a	Constant	-1.891	.419	-4.517	<.001	.151	.207	82.4 %
	ST-IAT	1.105	.416	2.658	.008	3.019		
1b	Constant	-1.847	.374	-4.942	<.001	.158	.327	81.4 %
	Explicit measure	1.300	.332	3.913	<.001	3.669		
2	Constant	-2.199	.516	-4.265	<.001	.111	.414	82.4 %
	ST-IAT	.842	.463	1.820	.069	2.322		
	Explicit measure	1.174	.399	2.939	.003	3.234		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting behavior (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

In sum, the results of Study 1 support findings and claims made in previous research on the predictive value of implicit attitudes. First, they suggest that explicit attitudes are better predictors for decided voters than implicit attitudes. In three out of four analyses, explicit attitudes were – on their own – better predictors of voting intention among decided respondents or voting behavior among those who had already voted. Second, and in line with the expectations made in previous research (Friese et al. 2012, 2016), implicit-explicit consistency was considerably lower than in studies dealing with implicit attitudes toward political parties or camps. Nevertheless, at least for the minimum wage initiative, implicit-explicit correspondence (Decided: r = -.37, Already voted: r = -.48) proved still too high for the implicit measure to provide predictive validity over and above the explicit measure.

One major limitation of Study 1 was the use of a WebInquisit ST-IAT which required participants to download a software plug-in. Unsurprisingly, downloading software raises concerns about malware and computer viruses among participants and thus acts as an effective impediment to data collection. In fact, the response rate in Study 1 was dramatically low (5.4 %). For this reason, I used a computer-administered paper-format ST-IAT in Study 2 which did not require respondents to download additional software.

2.4 Study 2

Between September 22, 2014 and September 27, 2014, I ran a second study using a computer-administered paper-format ST-IAT to measure participants' implicit attitudes toward the initiative "for a single health insurer." The initiative was launched by the Social Democratic Party (SP) and aimed at replacing the competitive system of health insurance by a single public health insurer. Voting took place on September 28, 2014. The initiative was rejected by a 61.9 percent majority.

2.4.1 Ethics statement

The same declaration applies as in the first study.

2.4.2 Participants

351 (42.7 % males, 57.3 % females) students, doctoral students, and employees of the University of Zurich volunteered for this study. Due to the different format of the ST-IAT, and thus the lack of appropriate information, no power analysis was conducted to determine sample size. However, since paper-format IATs tend to produce weaker IAT effects (Lemm et al. 2008), I made sure that sample size was larger than in Study 1. Participants were told that the study was about the upcoming federal votes and asked to take the survey only if they were eligible to vote. Their mean age was 26.79 years (SD=9.30).

2.4.3 Procedure

The Qualtrics (www.qualtrics.com) survey software was used to create the study. Participants answered the same questionnaires as in Study 1, along with two ST-IATs: a practice ST-IAT for implicit attitudes toward animals, and an ST-IAT for assessing their implicit attitudes toward the initiative for public health insurance.

2.4.4 Measures

Given the goal of overcoming the software plug-in impediment described above, I created an ST-IAT in line with Lemm et al.'s (2008) suggestions. Acknowledging the fact that computer-format IATs are sometimes not feasible, Lemm and colleagues (2008) developed an IAT that can be administered with paper and pencil and that mimics the results and psychometric properties of computer-based counterparts. In addition, they tested various scoring algorithms and made recommendations as to which algorithm should be used.

Both traditional computer IATs and paper-format IATs build on the assumption that the degree of congruence between a task and the implicit attitude of an individual determines the ease with which an individual can solve the task. In other words, if the pairing of a target with some evaluative category (e.g., positive) corresponds to the true implicit attitude of the

individual, it will be easier for her to sort stimuli in this task than when target and evaluative category are reversed. However, unlike traditional computer IATs, paper-format IATs do not measure response latencies. Instead, they measure the number of correct categorizations within a given time period. For example, in Lemm et al.'s (2008) studies, subjects were asked to categorize as many items as possible within a timeframe of 20 seconds. They were given a sheet with two columns of 20 items (i.e., trials) and asked to categorize items as either black or white names and pleasant or unpleasant words by marking the corresponding circles to the left and to the right of each item. After participants had completed the first condition, they received another 20 seconds for completing the second condition (in which category pairings were switched).

Building on the work of Lemm and colleagues (2008), I created a computer-administered paper-format ST-IAT for assessing participants' implicit attitude toward the initiative for public health insurance. As recommended by Lemm et. al (2008), subjects first completed a training ST-IAT for assessing their implicit attitude toward animals. Both training and critical ST-IAT consisted of six individual pages.

1. Participants were instructed to sort positive items by clicking the circle to the left of the item, while sorting negative items and items pertaining to the target category by clicking the right circle.

2. Once they felt ready to start, they could click on "Continue." On the next page, a countdown of five seconds appeared along with the message "The task will start in 5 seconds! Make yourself ready!"

3. Subjects categorized as many items as possible from a list of 25 items within a fixed time period of 20 seconds. A countdown timer was visible at the top of the page. After time had run out, the survey automatically advanced to the next page.

4. Instructions for the second combined block (Positive or Target vs. Negative) were provided.

5. Before starting the task, subjects were again given five seconds to prepare.

6. Participants categorized as many items as possible from a list of 25 items, again within 20 seconds.

As in the previous studies, block order was kept constant such that each participant first received the positive versus negative or referendum condition (see Appendix 2.B). In contrast to the previously administered ST-IAT versions, but in line with Lemm et al.'s (2008) suggestions, only word stimuli were used (see Appendix 2.C). Testing both paper-format IATs with verbal and picture stimuli and paper-format IATs with only verbal stimuli, Lemm and colleagues (2008) found stronger correlations between the latter and traditional computer IATs. Moreover, test-retest reliability was better in the case of paper-format IATs using only

verbal stimuli. For the evaluative categories, the same stimuli were used as in Study 1. The target category was represented by the names of two advocates (e.g., Jacqueline Fehr) of the initiative and the names of three parties (e.g., Green Party) supporting the initiative. ST-IAT scores were calculated using Lemm et al.'s (2008) *product: square root of difference (PSQoD)* approach. Comparing seven different algorithms, they find that *PSQoD* is most consistent with results from computer-format IATs. If A denotes the number of correct responses in the first block and B the number of correct responses in the second block, then *PSQoD* is calculated as...

$$(X/Y) \cdot \sqrt{X - Y},$$

where X is the greater of A or B, and Y is the smaller of A or B. Similar to Lemm et al. (2008), only participants with at least six responses in both blocks were retained for data analyses. This resulted in the exclusion of 10 participants and thus a sample of 341 participants. As can be seen above, *PSQoD* takes both the difference between the number of correct responses ($X - Y$) and the ratio of correct responses (X/Y) into account. In addition, it controls for extreme scores by taking the square root of the difference between X and Y. Note that if an equal number of correct categorizations are made in both blocks, the resulting ST-IAT score will equal zero and thus indicate indifferent implicit attitudes. Finally, in order to retain the directionality of the ST-IAT effect, the resulting values were multiplied by -1 if A was greater than B (thus indicating a negative attitude toward the referendum of interest).

Explicit attitudes were again measured with participants' self placement on the left-right scale. The explicit measure was strongly correlated with voting for the public health insurance initiative (r = -.56, p <.001).

Voting intention and voting behavior were measured using the same questions as in Study 1. Undecided voters (N=17) were again excluded from the analyses. This left 93 decided participants for the analysis of voting intention. Furthermore, there were no abstentions among those who had already voted on the referendum (N=231).

2.4.5 Results and discussion

Preliminary analyses

Because paper (ST-)IATs do not rely on trial by trial measurement of response latencies, they do not allow for the calculation of split-half or alpha reliability. I can thus not report these statistics. Implicit-explicit consistency was moderate (r = -.38) and thus very similar to the correlations found in the first study. As in the previous study, implicit-explicit correspondence was more pronounced among participants who had already voted on the initiative (r = -.43) than among those who simply reported their voting intention (r = -.32).

Voting intention and voting behavior

Tables 2.4 and 2.5 report the results from the logistic regression analyses for the prediction

of voting intention and voting behavior. As in the first study, the implicit attitude measure predicted voting intention and vote choice, yet it was a worse predictor than participants' self-reported (i.e., explicit) left-right placement. The explicit measure predicted 73.9 percent of decided voters correctly (Table 2.4) while the implicit measure could only predict 57.0 percent of cases. Similarly, implicit attitudes classified 66.5 percent of those respondents who had already voted correctly, while explicit attitudes allowed for the correct classification of 73.9 percent of cases. Thus, as in the previous study, explicit attitudes were a better predictor for decided voters.

Finally, implicit attitudes did not improve model fit in the case of voting intention. Adding the implicit measure to the explicit measure (model 2) only resulted in a slight increase in Nagelkerke's R^2 (+1.9 percentage points) and no change in the %CCC. Finally, the implicit measure did not substantially improve the quality of the prediction of voting behavior (+0.3%CCC) once explicit attitudes were controlled for.

Table 2.4: Results of logistic regression for prediction of voting intention (Study 2)

Step	**Variable**	**B**	***SE***	**Wald**	***p***	**Exp(B)**	**Nagelkerke's R^2**	**% CCC**
Public health insurance initiative (N=93)								
1a	Constant	-.104	.216	-.480	.631	.901	.102	57.0 %
	ST-IAT	.609	.242	2.519	.012	1.838		
1b	Constant	-.169	.257	-.658	.511	.845	.393	73.9 %
	Explicit measure	-1.506	.343	-4.395	<.001	.222		
2	Constant	-.164	.260	-.630	.528	.849	.412	73.9 %
	ST-IAT	.369	.275	1.340	.180	1.446		
	Explicit measure	-1.423	.346	-4.116	<.001	.241		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting intention (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

Table 2.5: Results of logistic regression for prediction of voting behavior (Study 2)

Step	Variable	B	*SE*	Wald	*p*	Exp(B)	Nagelkerke's R^2	% CCC
Public health insurance initiative (N=230)								
1a	Constant	.263	.141	1.866	.062	1.301	.149	66.5 %
	ST-IAT	.756	.158	4.788	<.001	2.130		
1b	Constant	.239	.161	1.490	.136	1.270	.395	73.9 %
	Explicit measure	-1.471	.202	-7.290	<.001	.230		
2	Constant	.242	.162	1.492	.136	1.274	.406	74.2 %
	ST-IAT	.326	.186	1.755	.079	1.386		
	Explicit measure	-1.337	.209	-6.394	<.001	.263		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting behavior (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

To summarize, both studies suggest that implicit attitudes predict voting intention and voting behavior in Swiss referendums. However, at least for decided voters, they are worse predictors than explicit measures such as self-reported left-right placement. More importantly, although the correlations between the implicit and explicit measure were generally lower than in studies on elections (e.g., Friese et al. 2012), implicit attitudes did only provide little incremental validity (between 0.3 and 1.0 percentage points) and in one case (i.e., the minimum wage initiative) they even decreased the %CCC. However, studies 1 and 2 were limited in scope to decided voters. Hence, they could not address the claim made in previous research (e.g., Galdi et al. 2008) that implicit attitudes are better predictors for undecided voters as compared to decided voters. To evaluate this claim, I conducted an additional study. Because of the more pragmatic (i.e., more user-friendly) character of the computer-administered paper-format ST-IAT, I used this particular test version in a third study combining both a pre- and post-vote survey.

2.5 Study 3

Study 3 took place three weeks ahead of the national vote (November 30, 2014) on the so-called "Ecopop initiative." Launched by the organization Ecology and Population (ECOPOP), the referendum proposed limiting the annual net migration to 0.2 percent of Switzerland's resident population. Unlike most other anti-immigration referendums, the Ecopop initiative did not refer to immigration as a threat to Swiss society or its economy but rather as a major cause of environmental degradation. It was thus believed to attract votes from both right-wing voters and, to a lesser degree, voters identifying with the Green and Alternative Left Parties. Due to its likely negative impact on the Swiss economy and its xenophobic character, the referendum faced opposition from all major political parties and the Swiss government. It was eventually

rejected by a clear majority of 74.1 percent.

2.5.1 Ethics statement

The study was conducted online. All subjects were older than 18 years of age. Data collection was performed by respondi AG (www.respondi.de) which conforms to the ESOMAR codes and guidelines for online access panels. On the first page, subjects received brief information on the purpose of the study, the measures involved, and the leading research institute (Department of Political Science, University of Zurich). At the end of the study, participants received contact information. Debriefing was provided on request. All data was analyzed anonymously and no identifying data (e.g., names) were collected. At the author's faculty (Faculty of Arts, University of Zurich) the IRB asks researchers to self-assess the ethical soundness of their research using a checklist and only seek approval from the IRB if one of the questions is answered affirmatively. This was not the case for the present study.

2.5.2 Participants

Sampling was done using the online access panel provided by respondi AG. Only eligible voters (N=1824) from the German- and French-speaking parts of Switzerland were recruited. Retaining participants with at least six correct responses in the ST-IAT blocks resulted in the exclusion of 183 participants. Of the remaining 1641 participants, 1006 participants (37.48 % males, 62.52 % females) reported that they had not yet voted and 862 participants (38.98 % males, 61.02 % females) answered affirmatively when asked whether they intended to vote. Of those, 704 returned for the second survey, which took place in the week after the vote (return rate of 81.67 %, 41.05 % males, 58.95 % females), and 552 eventually reported their vote choice (44.38 % males, 55.62 % females). Mean age among participants in the final sample was 43.35 years (SD=13.87). Participants received a small compensation of EUR 1.10 for participation in the first survey and EUR 0.50 for participation in the second survey. The final sample consisted of 457 participants who initially reported to be decided and 82 participants who reported to be undecided. Due to the low return rate among undecided voters (49.10 %), the final sample of undecided voters was slightly smaller than the one determined a priori using G*Power (Faul et al. 2009), a stand-alone power analysis program for statistical tests. The analysis indicated that for the desired power (.80), alpha level (.05), anticipated (i.e., based on Study 2) effect size (OR=2.0), and two moderately correlated predictors (R^2 other X = 0.15), a minimum sample size of 97 was required.

2.5.3 Procedure

In the pre-vote survey, participants completed several questionnaires on socio-demographic variables, political variables (e.g., partisanship, left-right placement), and their concern about the impact of immigration on the economy, culture, and environment. At the end of the survey, they completed the same practice ST-IAT used in the previous study, followed by another

ST-IAT for assessing their implicit attitude toward "Ecopop." After voting day, subjects received an invitation to take the second survey, in which they were asked to report their vote choice. In addition, several questions serving as proxy variables for cognitive elaboration were included. Both surveys were programmed in Qualtrics.

2.5.4 Measures

Implicit attitudes were measured using the same ST-IAT as in the previous study. For the target category (i.e., Ecopop initiative), I used the names of three politicians, one right-wing party, and one political association supporting the referendum. In order to account for differences between language regions, I used slightly different stimuli in the French version of the ST-IAT than in the German ST-IAT (see Appendix 2.D).

Explicit attitudes were assessed with three items. Participants were asked to indicate their concern about the impact of immigration on the (i) economy, (ii) Swiss culture, and (iii) environment using 6-point Likert items (1 = *very unconcerned* to 6 = *very concerned*) (α = 0.86). Items were averaged to form the concern scale, with high numbers reflecting higher concern about the impact of immigration. Voting on Ecopop was strongly associated (r = .49, p <.001) with explicit attitudes.

Cognitive elaboration was measured with two items ($\alpha = 0.59$) that were exclusively created for the purpose of this study. First, participants were asked to indicate whether they perceived the issue of the referendum to be an old or novel political issue using a 6-point Likert item (1 = *completely novel* to 6 = *completely old*). Second, participants reported whether and how often they had already voted on topically similar referendums (1 = *never or very rarely* to 6 = *very often*). Both items were averaged to create the cognitive elaboration index with higher values indicating more elaborated attitudes toward immigration and the referendum of interest. Although it seems somewhat counterintuitive, there was no relationship between the age of participants and their cognitive elaboration ($r = 0.07$, ns).

Voting intention and voting behavior were measured in the same way as in Studies 1 and 2. In the pre-vote survey, 692 participants reported their vote intention and 167 were undecided. In the second survey (i.e., after the vote), 13 of 552 participants reported that they had abstained from the vote. This left a total of 539 participants for the analyses of voting behavior.

2.5.5 Results and discussion

Preliminary analyses

I first tested for consistency between the implicit and explicit measure for both decided and undecided voters. Overall, correspondence between the implicit and explicit measure was lower (r = .23) than in the previous studies. In addition, the two types of measures were only correlated among decided voters (r = .27) but not among undecided voters (r = .01). Finally, regressing the explicit measure on the implicit measure, cognitive elaboration and their

interaction revealed that implicit-explicit attitude correspondence was marginally ($\beta = .088$, $t(539) = 1.874$, $p = .062$) stronger at higher levels of cognitive elaboration. The results of this analysis are included in the Appendix 2.E. These results are consistent with previous research (e.g., Friese et al. 2012) arguing that consistency between implicit and explicit attitudes is more pronounced among decided voters as a result of their greater cognitive elaboration of attitudes. In sum, implicit-explicit correspondence was relatively low and closer to the correlations found in Galdi et al. (2008) than in pre-election studies (e.g., Friese et al. 2012).

Voting intention

On their own, both the implicit (model 1a) and explicit (model 1b) measure predicted voting intention (see Table 2.6). However, as in the previous studies, the explicit measure better predicted voting intention. Moreover, looking at the incremental validity of implicit attitudes, I found that the latter increased Nagelkerke's R^2 by 1.8 percentage points after controlling for explicit attitudes (model 2) but decreased the percentage of correctly classified cases by 0.4 percentage points. On the other side, the explicit measure increased Nagelkerke's R^2 by 29.0 percentage points after accounting for implicit attitudes, while correctly classified cases increased by 8.2 percentage points. In sum, the explicit measure performed better in terms of both predictive validity in general and incremental validity.

Table 2.6: Results of logistic regression for prediction of voting intention (Study 3)

Step	**Variable**	**B**	***SE***	**Wald**	***p***	**Exp(B)**	**Nagelkerke's R^2**	**% CCC**
Ecopop initiative (N=691)								
1a	Constant	-.607	.083	-7.343	<.001	.545	.090	66.0 %
	ST-IAT	.572	.088	6.467	<.001	1.771		
1b	Constant	-.835	.101	11.716	<.001	.434	.362	74.6 %
	Explicit measure	1.460	.125	-8.296	<.001	4.307		
2	Constant	-.844	.102	-8.187	<.001	.430	.380	74.2 %
	ST-IAT	.357	.102	3.506	<.001	1.429		
	Explicit measure	1.378	.126	10.943	<.001	3.969		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting intention (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

Voting behavior

Table 2.7 below presents results from binary logistic regression models for the prediction of voting behavior among decided and undecided voters. Implicit (model 1a) and explicit (model 1b) attitudes both predicted voting behavior of decided and undecided voters. However, a comparison of the implicit-only models (models 1a) for decided and undecided voters reveals that, on their own, implicit attitudes are better predictors of voting behavior among undecided voters, with 78.1 percent of them being correctly classified (as compared to 73.0 percent among decided voters). In line with these descriptive results, implicit attitudes improved model fit more for undecided voters than for decided voters once explicit attitudes were controlled for. In fact, Nagelkerke's R^2 increased by only 1.1 percentage points for decided voters, while it increased by 5.5 percentage points for undecided voters upon adding the implicit measure to the explicit measure (model 2). However, in regards to the percentage of correctly classified cases, implicit attitudes improved model fit to an equal extent for decided (+1.1%CCC) and undecided (+1.2%CCC) voters. On the other side, explicit attitudes increased Nagelkerke's R^2 by 32.0 percentage points and 13.0 percentage points for decided and undecided voters after accounting for implicit attitudes, while correctly classified cases increased by 6.8 and 0.9 percentage points for decided and undecided voters. Taken together, from a descriptive point of view, explicit attitudes better predicted vote choice of decided voters, while implicit attitudes fared slightly better than explicit attitudes in their prediction of vote choice among undecided voters.

Table 2.7: Results of logistic regression for prediction of voting behavior, separately for decided and undecided voters (Study 3)

Step	Variable	B	*SE*	Wald	*p*	Exp(B)	Nagelkerke's R^2	% CCC
Decided voters who voted (N=457)								
1a	Constant	-1.003	.110	-9.127	<.001	.367	.081	73.0 %
	ST-IAT	.593	.121	4.899	<.001	1.809		
1b	Constant	-1.380	.146	-9.421	<.001	.252	.390	78.7 %
	Explicit measure	1.574	.166	9.511	<.001	4.827		
2	Constant	-1.390	.148	-9.410	<.001	.249	.401	79.8 %
	ST-IAT	.335	.139	2.408	.016	1.399		
	Explicit measure	1.491	.168	8.887	<.001	4.440		
Undecided voters who voted (N=82)								
1a	Constant	-1.378	.292	-4.727	<.001	.252	.082	78.1 %
	ST-IAT	.644	.322	1.997	.046	1.904		
1b	Constant	-1.514	.323	-4.694	<.001	.220	.157	77.8 %
	Explicit measure	.977	.360	2.712	.007	2.657		
2	Constant	-1.635	.353	-4.628	<.001	.195	.212	79.0 %
	ST-IAT	.619	.358	1.731	.083	1.857		
	Explicit measure	.948	.367	2.580	.010	2.580		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting behavior (0=No, 1=Yes). All continuous variables were z-standardized separately for decided and undecided voters prior to the analyses.

To complement these descriptive analyses, I conducted moderation analyses to further examine the potential interaction between the implicit and explicit measure and decidedness. Table 2.8 shows that, when modeled separately, neither the explicit (model 1) nor the implicit measure (model 2) predicted voting behavior significantly better for decided individuals. Although both coefficients point in the expected direction, they do not reach conventional levels of significance. The same can be said when both implicit and explicit attitudes are considered along with decidedness and their respective interactions (model 3). Thus, implicit attitudes predicted voting behavior descriptively, but not significantly better for undecided voters while explicit attitudes predicted voting behavior descriptively, but not significantly better for decided respondents.

Table 2.8: Results of logistic regression analyses predicting voting behavior from explicit (EA) and implicit attitudes (IA) and decidedness (Study 3)

Model	Variable	B	*SE*	Wald	*p*	Exp(B)	Nagelkerke's R^2	% CCC
1	Explicit attitudes (N=532)							
	Constant	-1.581	.335	-4.719	<.001	.206	.361	78.6 %
	EA	1.101	.406	2.712	.007	3.007		
	Decidedness	.227	.365	.622	.534	1.255		
	EA*Decidedness	.512	.440	1.164	.245	1.669		
2	Implicit attitudes (N=538)							
	Constant	-1.452	.306	-4.744	<.001	.234	.084	73.8 %
	IA	.619	.310	1.997	.046	1.858		
	Decidedness	.437	.325	1.342	.180	1.548		
	IA*Decidedness	-.020	.333	-.058	.953	.981		
3	Explicit and implicit attitudes (N=531)							
	Constant	-1.771	.382	-4.638	<.001	.170	.377	79.7 %
	EA	1.068	.414	2.580	.010	2.910		
	IA	.595	.344	1.731	.083	1.814		
	Decidedness	.399	.409	.974	.330	1.490		
	EA*Decidedness	.459	.448	1.025	.305	1.583		
	IA*Decidedness	-.256	.372	-.689	.491	.774		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting behavior (0=No, 1=Yes). All continuous variables were z-standardized prior to the analyses.

Does cognitive elaboration moderate the influence of implicit attitudes?

To explore whether the cognitive elaboration of attitudes moderated the predictive validity of the implicit and explicit measure, I ran logistic regression analyses in which voting behavior was modeled as a function of the implicit or explicit measure, cognitive elaboration and their interaction. Table 2.9 shows that cognitive elaboration moderated the influence of the explicit (model 1) and implicit measure (model 2) on voting behavior, but only with marginal significance. Furthermore, when modeled together (model 3), these interactions became non-significant. However, separate estimations at one standard deviation above and below the mean of cognitive elaboration revealed that, at least for implicit attitudes, predictive validity was better for voters with well-elaborated attitudes toward the referendum (see Figure 2.1). That is, as elaboration increased, the influence of the ST-IAT on voting behavior became more pronounced.

Table 2.9: Results of multiple binary logistic regression analyses for the prediction of voting behavior including cognitive elaboration as moderator (Study 3)

Model	Variable	B	*SE*	Wald	*p*	Exp(B)	Nagelkerke's R^2	% CCC
1	Explicit attitudes (N=531)							
	Constant	-1.407	.136	-10.332	<.001	.245	.360	78.3 %
	EA	1.553	.159	9.748	<.001	4.728		
	Elaboration	-.183	.140	-1.314	.189	.833		
	EA*Elaboration	.260	.148	1.759	.079	1.297		
2	Implicit attitudes (N=537)							
	Constant	-1.086	.105	-10.341	<.001	.338	.092	74.3 %
	IA	.601	.115	5.242	<.001	1.825		
	Elaboration	-.185	.103	-1.788	.074	.832		
	IA*Elaboration	.218	.125	1.741	.082	1.244		
3	Explicit and implicit attitudes (N=530)							
	Constant	-1.442	.139	-10.374	<.001	.237	.375	79.1 %
	EA	1.461	.161	9.078	<.001	4.309		
	IA	.377	.131	2.865	.004	1.457		
	Elaboration	-.191	.142	-1.350	.177	.826		
	EA*Elaboration	.222	.149	1.482	.138	1.248		
	IA*Elaboration	.112	.147	.765	.444	1.119		

Notes: B = regression weight; *SE* = standard error of the regression weight; Wald = Wald criterion; Exp(B) = Odds ratio. Relative amount by which the odds increase (Exp(B)>1.0) or decrease (Exp(B)<1.0) when the value of the predictor is increased by 1 unit; CCC = correctly classified cases; DV = voting behavior (0=No, 1=Yes). EA = explicit attitudes; IA = implicit attitudes; All continuous variables were z-standardized prior to the analyses.

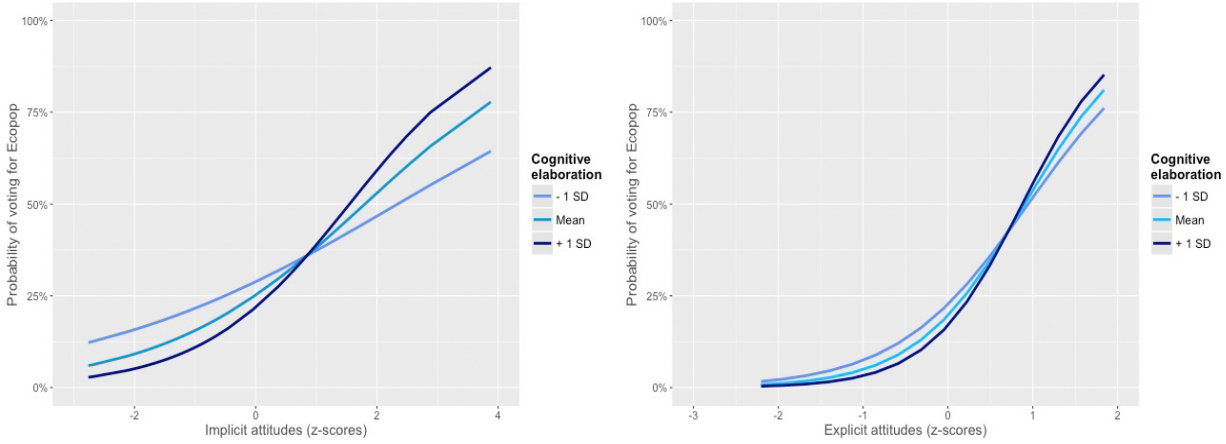

Figure 2.1: Probability of voting for Ecopop as a function of implicit attitudes, cognitive elaboration, and their interaction (left), and as a function of explicit attitudes, cognitive elaboration, and their interaction (right). The implicit measure (ST-IAT) predicted vote choice better for voters scoring high on cognitive elaboration (+1 SD) than voters scoring low on cognitive elaboration (−1 SD) as indicated by the steeper line for high cognitive elaboration.

2.6 General Discussion

The use of implicit attitude measures has been proposed to tackle the challenge of predicting the vote of undecided voters. Most of the optimism has stemmed from a seminal study (Galdi et al. 2008) in which implicit, but not explicit, attitudes were found to be significant predictors of undecided respondents' opinions on a political issue. Interestingly, the pattern of results was reversed for decided respondents: Explicit, but not implicit, attitudes predicted future opinions. Against the backdrop of these results, a full double dissociation pattern (Galdi et al. 2008; Gawronski and Galdi 2011) has been proposed comprising four hypotheses:

1. Implicit attitudes predict voting behavior better than explicit attitudes for undecided voters.

2. Explicit attitudes predict voting behavior better than implicit attitudes for decided voters.

3. Implicit attitudes predict voting behavior better for undecided than decided voters.

4. Explicit attitudes predict voting behavior better for decided than undecided voters.

However, recent studies have shed doubt on the incremental validity of implicit attitudes in general and their utility for undecided voters in particular. Reviewing the more recent evidence on the predictive value of implicit attitudes, Friese et al. (2016, 19) conclude that "[i]n most cases [...], the increase in %CCC remained well below 1 percentage point, sometimes there was no change at all, and sometimes even descriptively a decrease in %CCC occurred". What is more, the authors also mention (p. 19) that there is "[...] no evidence for the ideas that implicit measures predict the voting behavior of undecided voters (a) better than explicit measures do, or (b) better than they predict the behavior of decided voters". In an attempt to explain why replications of the double dissociation pattern have generally failed, Friese et al. (2012) point to the contextual differences between the seminal study by Galdi et al. (2008) and subsequent research. In fact, while the former dealt with implicit attitudes toward a local political issue where no actual voting took place, subsequent studies have focused on real political elections and voting behavior. In such a context, implicit and explicit measures tend to overlap as a result of well elaborated attitudes toward parties or candidates and this will, eventually, leave little room for implicit measures to predict beyond explicit measures.

It follows from the above reasoning that implicit measures may be better predictors in the context of voting on specific political issues where attitudes tend to be less elaborated. For this reason, I have tested the predictive validity of implicit attitudes in the context of issue-related votes in Switzerland. In the remainder of this section, I shall briefly discuss my results with respect to the four hypotheses presented above.

1. Implicit attitudes predict voting behavior better than explicit attitudes for undecided voters. Only the design of the third study allowed for a comparison between

decided and undecided voters. Unlike stated in the hypothesis, implicit and explicit attitudes were equally good predictors of vote choice among undecided voters. On their own, implicit attitudes correctly predicted the vote of 78.1 percent of undecided respondents while explicit attitudes did so for 77.8 percent.

2. Explicit attitudes predict voting behavior better than implicit attitudes for decided voters. All three studies support this claim. In Studies 1 and 2, self-reported left-right placement was a better predictor of participants' voting intention (minimum wage initiative: IA: 62.8 %CCC, EA: 81.8 %CCC; Gripen referendum: IA: 85.1 %CCC, EA: 85.3 %CCC; public health insurance initiative: IA: 57.0 %CCC, EA: 73.9 %CCC) and in Study 3 participants' concern about immigration better predicted voting behavior among decided voters (IA: 73.0 %CCC, EA: 78.7 %CCC).

3. Implicit attitudes predict voting behavior better for undecided than decided voters. In Study 3, I found some, albeit descriptive, evidence for this hypothesis. The implicit measure was a better predictor for the voting behavior of undecided (78.1 %CCC) than decided (73.0 %CCC) voters. However, implicit attitudes improved the quality of the overall prediction only slightly for both decided (+1.1%CCC) and undecided voters (+1.2%CCC) and there was no significant interaction between implicit attitudes and decidedness.

4. Explicit attitudes predict voting behavior better for decided than undecided voters. Consistent with this hypothesis, explicit attitudes better predicted voting behavior among decided voters (78.7 %CCC) than among undecided voters (77.8 %CCC). In addition, they improved model fit to a greater extent for decided voters (+6.8%CCC) than undecided voters (+0.9%CCC). Nevertheless, these descriptive analyses were not backed up by a subsequent moderation analysis.

Finally, in Study 3, I found some evidence in favor of the argument raised in previous research that greater elaboration of attitudes would lead to better predictive validity of implicit attitudes. It should be noted, however, that the conceptualization of cognitive elaboration in Study 3 was most likely underdeveloped. In fact, the two items that were used for the measurement of cognitive elaboration both pertained to the frequency with which participants had thought about the substantive issue of the vote. However, in previous research (e.g., Nosek 2005) on the moderators of implicit-explicit consistency, thought frequency was only one of three aspects of cognitive elaboration. In fact, importance and familiarity were also considered. Hence, it is likely that the interaction between implicit/explicit attitudes and elaboration was not fully picked up because of the incomplete conceptualization and measurement of cognitive elaboration.

Taken together, my results stand in-between those of Friese and colleagues (2012) and those of Galdi et al. (2008). As in the former study, explicit attitudes were better predictors of voting behavior among decided voters. However, unlike results of Friese et al. (2012) but similar to Galdi et al. (2008), implicit attitudes did, at least from a descriptive perspective, a better job in predicting choices of undecided voters as compared to decided voters. Yet, unlike in Galdi and colleagues study, they did not outperform explicit attitudes. Thus, neither the results of Galdi and colleagues nor the results of Friese and colleagues could be replicated.

Given that the correspondence between the implicit and explicit measure was fairly low in all three studies (r = .23 - .40), it is surprising that the double dissociation pattern described above could not be replicated. Researchers should, however, keep in mind that the cognitive elaboration of attitudes may not necessarily represent the only moderator of the predictive validity of implicit measures. For example, research on attitudinal ambivalence suggests that ambivalent individuals rely less on their implicit attitudes when making decisions. Hence, implicit attitudes should be less relevant for ambivalent, and thus undecided, voters. This example shows that a multitude of theoretical models and mechanisms exists explaining why and under what conditions implicit attitudes should predict voting behavior of undecided individuals. Researchers will need to rethink some of these theoretical accounts while, at the same time, test others before implicit attitudes will allow for considerable improvements of political predictions.

References

Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., and Alessandro Amadori. 2008. Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, 29(3), 369-387.

Bluemke, M., and Malte Friese. 2008. Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, 38, 977-997.

Dotsch, R., Wigboldus, D.H.J., and Ad van Knippenberg. 2011. Biased Allocation of Faces to Social Categories. *Journal of Personality and Social Psychology*, 100(6), 999-1014.

Dovidio, J.F., Kawakami, K., Johnson, C., Johnson B., and Adaiah Howard. 1997. On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, 33(5), 510-540.

Faul, F., Erdfelder, E., Buchner, A., and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavioral Research Methods*, 41: 1149-1160.

Friese, M., Bluemke, M., and Michaela Wänke. 2007. Predicting Voting Behavior with Implicit Attitude Measures. *Experimental Psychology*, 54(4), 247-255.

Friese, M., Smith, C.T., Plischke, T., Bluemke, M., and Brian A. Nosek. 2012. Do implicit attitudes predict voting behavior particularly for undecided voters? *PLoS ONE*, 7(8).

Friese, M., Smith, C.T., Koever, M., and Matthias Bluemke. 2016. Implicit Measures of Attitudes and Political Voting Behavior. *Social and Personality Psychology Compass*, 10(4), 188-201.

Galdi, S., Arcuri, L., and Bertram Gawronski. 2008. Automatic Mental Associations Predict Future Choices of Undecided Decision-Makers. *Science*, 321, 1100-1102.

Galdi, S., Gawronski, B., Arcuri, L., and Malte Friese. 2012. Selective exposure in decided and undecided individuals: Differential relations to automatic associations and conscious beliefs. *Personality and Social Psychology Bulletin*, 38, 559-569.

Gawronski, B., Hofmann, W., and Christopher J. Wilbur. 2006. Are "implicit" attitudes unconscious? *Consciousness and Cognition*, 15(3), 485-499.

Gawronski, B., and Silvia Galdi. 2011. Using implicit measures to read the minds of undecided voters. In: Cadinu, M., Galdi, S., and Anne Maass (Eds.), *Social perception, cognition, and language in honour of Arcuri*, 203-216. Padova: CLEUP.

Greenwald, A.G., McGhee, D.E., and Jordan L.K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.

Greenwald, A.G., Nosek, B.A., and Mahzarin R. Banaji. 2003. Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, 85(2), 197-216.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., and Mahzarin R. Banaji. 2009. Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.

Karpinski, A., Steinman, R.B., and James L. Hilton. 2005. Attitude Importance as a Moderator of the Relationship Between Implicit and Explicit Attitude Measures. *Personality and Social Psychology Bulletin*, 31(7), 949-962.

Karpinski, A., and Ross B. Steinman. 2006. The single category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, 91, 16-32.

Lemm, K.M., Lane, K.A., Sattler, D.N., Khan, S.R., and Brian A. Nosek. 2008. Assessing Implicit Cognitions with a Paper-Format Implicit Association Test. In: Morrison, M.A., and Todd G. Morrison (Eds.), *The Psychology of Modern Prejudice*, 123-146. New York: Nova Science Publishers.

Lundberg, K.B., and Keith Payne. 2014. Decisions among the Undecided: Implicit Attitudes Predict Future Voting Behavior of Undecided Voters. *PLoS ONE*, 9(1).

Nosek, Brian A. 2005. Moderators of the Relationship between Implicit and Explicit Evaluation. *Journal of Experimental Psychology General*, 134, 565-584.

Nosek, B.A., Graham, J., and Carlee B. Hawkins. 2010. Implicit Political Cognition. In: Gawronski, B., and B.K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*, 548-564. New York: The Guilford Press.

Perugini, Marco. 2005. Predictive Models of Implicit and Explicit Attitudes. *British Journal of Social Psychology*, 44, 29-45.

Roccato, M., and Cristina Zogmaister. 2010. Predicting the Vote through Implicit and Explicit Attitudes: A Field Research. *Political Psychology*, 31(2), 249-274.

Visser, P.S., Krosnick, J.A., Marquette, J., and Michael F. Curtin.2000. Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters, and Response Order Effects. In: Lavrakas, P., and M. Traugott (Eds.), *Election Polls, the News Media, and Democracy*, 224-260. New York: Chatham House.

Appendix

2.A Stimuli List for ST-IATs on Minimum Wage Initiative & Gripen Referendum (Study 1)

Table 2.A.1: Stimuli List for ST-IATs on Minimum Wage Initiative & Gripen Referendum (Study 1)

Category	**Stimuli**
positive	Love (Liebe), Joy (Freude), Paradise (Paradies), Gift (Geschenk), Holiday (Ferien)
negative	Poison (Gift), Stink (Gestank), Disease (Krankheit), Disaster (Katastrophe), Death (Tod)
Minimum Wage Initiative	Cédric Wermuth (words), Christian Levrat (words), Campaign Poster (picture), SP party emblem, UNIA emblem
Gripen Referendum	Schweizer Armee (words), Ueli Maurer (words), Campaign Poster (picture), Gripen fighter jet (picture), André Blattmann (photo)

Note: Original German positive and negative stimuli are in brackets. Target stimuli were words and pictures.

2.B First condition of computer-administered paper-format ST-IAT (Study 2)

Table 2.B.1: First condition of computer-administered paper-format ST-IAT (Study 2)

Positive	Negative
	or
	Public Health Insurance Initiative

○	Love	○
○	Joy	○
○	Jacqueline Fehr (SP)	○
○	SP	○
○	Paradise	○
○	Yvonne Gilli (Green)	○
○	Poison	○
○	Gift	○
○	Holidays	○
○	Green Party	○
○	EVP	○
○	Love	○
○	Stink	○
○	Joy	○
○	Jacqueline Fehr (SP)	○
○	SP	○
○	Disease	○
○	Disaster	○
○	Holidays	○
○	Gift	○
○	EVP	○
○	Paradise	○
○	Death	○
○	Green Party	○
○	Yvonne Gilli (Green)	○

2.C Stimuli List for ST-IAT on Public Health Insurance Initiative (Study 2)

Table 2.C.1: Stimuli List for ST-IAT on Public Health Insurance Initiative (Study 2)

Category	Stimuli
positive	Love (Liebe), Joy (Freude), Paradise (Paradies), Gift (Geschenk), Holiday (Ferien)
negative	Poison (Gift), Stink (Gestank), Disease (Krankheit), Disaster (Katastrophe), Death (Tod)
Public Health Insurance Initiative	Jacqueline Fehr (SP), SP, Yvonne Gilli (Grüne), Grüne Partei, EVP

Note: Original German positive and negative stimuli are in brackets. Target stimuli were only words.

2.D Stimuli List for ST-IAT on Ecopop Initiative (Study 3)

Table 2.D.1: Stimuli List for ST-IAT on Ecopop Initiative (Study 3)

Category	Stimuli
positive	Love (Liebe/Amour), Joy (Freude/Joie), Paradise (Paradies/Paradis), Gift (Geschenk/Cadeau), Holiday (Ferien/Vacances)
negative	Poison (Gift/Poison), Stink (Gestank/Puanteur), Disease (Krankheit/Maladie), Disaster (Katastrophe/Catastrophe), Death (Tod/Mort)
Ecopop Initiative (German)	Thomas Minder, Schweizer Demokraten (SD), Hans Geiger (SVP), Aktion für eine Unabhängige und Neutrale Schweiz (AUNS), Pirmin Schwander (SVP)
Ecopop Initiative (French)	Thomas Minder, Démocrates Suisses (DS), Philippe Roch, Action pour une Suisse Indépandante et Neutre (ASIN), Pirmin Schwander (SVP)

Note: Original German and French positive and negative stimuli are in brackets. Target stimuli were only words.

2.E Results of linear regression of the explicit measure on the implicit measure, cognitive elaboration, and their interaction

Table 2.E.1: Results of linear regression of the explicit measure on the implicit measure, cognitive elaboration, and their interaction

Variable	**B**	***SE***	**t**	***p***	**R^2**	**Adj. R^2**	**F test**
Constant	-.035	.040	-.862	.389	.080	.074	15.537
ST-IAT	.243	.042	5.800	<.001			
Elaboration	-.115	.040	-2.867	.004			
ST-IAT*Elaboration	.088	.047	1.874	.062			

Notes: N = 543; B = regression weight; *SE* = standard error of the regression weight; t = t value; DV = explicit measure (i.e., concern about the impact of immigration). All continuous variables were z-standardized prior to the analysis.

3

The Causal Mediation Effect of Likeability on Female Students' Career Prospects in Science

Livio Raccuia and Marco Steenbergen

WOMEN are still underrepresented in academic science. Recent studies suggest that a great deal of these imbalances are due to gender discrimination resulting from implicit gender stereotypes. We replicated previous research on the detrimental effects of female student gender on factors relevant to careers in science using the newly-proposed parallel encouragement design (PED) and Amazon Mechanical Turk (MTurk). In contrast to previous research, we found that although MTurk subjects exhibited implicit gender stereotypes, they still favored female engineering students by granting them more starting salary. Our experimental design further showed that, other than we expected, the effect was not mediated by perceived competence but likeability of the target student. That is, MTurk subjects favored female engineering students because they viewed them as more likeable than their male counterparts. The same causal mechanism neutralized an otherwise negative effect of female student gender on subjects' perceptions of the hireability of the target student. In contrast to what we expected, subjects' implicit gender bias did not moderate these mediation effects. Our results thus suggest that participants did not apply their implicit gender stereotypes when judging the aptitude of engineering students and making salary offers.

3.1 Introduction

Over the last decade, social psychology has witnessed a proliferation of research on implicit stereotypes. Typically defined as the automatic association of a trait with a particular social group or category (Kang 2009), implicit stereotypes were shown to be both pervasive and robust (Greenwald et al. 1998; Kang and Lane 2010; Nosek et al. 2007). Acquired very early in life as a result of personal experience, observation, and media exposure (Eagly 1987; Staats 2014) implicit stereotypes may later lead to biased (i.e., unobjective) judgments and behaviors if the traits that are associated with a particular social group are predominantly positive or negative. While most of the initial research had focused on the detrimental effects of implicit race bias on attitudes and behavior toward Black Americans (e.g., Amodio and Devine 2006; Correll et al. 2002; Dovidio and Gaertner 2000; Ziegert and Hanges 2005) more recent research has investigated the prevalence and impact of implicit gender stereotypes. In fact, the pervasiveness of implicit gender stereotypes is well documented. For example, in a large-scale project run by the University of Harvard (Nosek et al. 2002), the dominant implicit association among respondents was for female with family and male with career rather than vice versa. Similar results were found in a gender-science implicit association test (IAT) with most subjects strongly associating female with liberal arts rather than science (Nosek et al. 2002).

More importantly, stereotypes of this sort have been claimed responsible for the discrimination of women in the hiring and employment process. In two very recent studies, both Moss-Racusin et al. (2012) and Reuben et al. (2014) examined the effect of gender and subtle (Moss-Racusin et al. 2012) or implicit gender bias (Reuben et al. 2014) on hireability in academic science. In both experiments, women were less likely to be hired for scientific positions or tasks either because they were viewed as less competent (Moss-Racusin et al. 2012) or as a direct function of implicit stereotypes (Reuben et al. 2014). In addition, the negative effect of female gender identity on perceived competence, hireability, and salary conferral was found to be more pronounced for subjects high on subtle gender bias (Moss-Racusin et al. 2012). While we applaud and appreciate the efforts of both groups of scholars to incorporate implicit gender bias in their accounts of gender-based discrimination, we nonetheless believe that both studies have considerable limitations. First, we believe that reporting direct effects of implicit gender stereotypes on women's hireability is too simplistic as it does not capture the true role of implicit gender stereotypes and thus obscures the underlying causal mechanism. Hence, we deem research that includes implicit gender bias as a third variable mediating or moderating the effect of female gender as a more appropriate approach. In fact, Moss-Racusin et al. (2012) found that female students were perceived as less hireable because science faculty viewed them as less competent than male student applicants. However, and second, their mediation analysis followed recommendations by Baron and Kenny (1986), a method that was recently shown to be outdated because it ignores the assumptions that are necessary for the identification of causal mediation effects (Imai et al. 2010a). Finally,

Moss-Racusin and colleagues (2012) measured subtle gender bias using the modern sexism scale (Swim et al. 1995), a measure they consider to "tap a subtle, modern form of gender bias that often exists outside of individuals' conscious awareness." While we agree that it is unconscious or implicit bias rather than overt sexism that accounts for gender discrimination in academic science, we doubt that the modern sexism scale is a non-reactive measure as the authors argue. Although the modern sexism scale was shown to be uncorrelated with traditional measures of social desirability (Campbell et al. 1997), we deemed it a better approach to measure implicit gender bias by having subjects complete an implicit association test (IAT; Greenwald et al. 1998). After all, IATs have become the most widely used implicit measure and they have been shown to be less reactive than explicit measures, particulary in socially sensitive domains (Greenwald et al. 2009). For all these reasons, we chose to conduct a replication of the Moss-Racusin et al. (2012) study using the parallel encouragement design (PED) (Imai et al. 2011) and measuring implicit gender bias by means of an IAT. Based on the original study, we expected a negative effect of female gender identity on outcomes such as perceived hireability and salary conferral that would be mediated by perceived competence. Furthermore, we expected that the effect would be moderated by implicit gender stereotypes (see Figure 3.1).

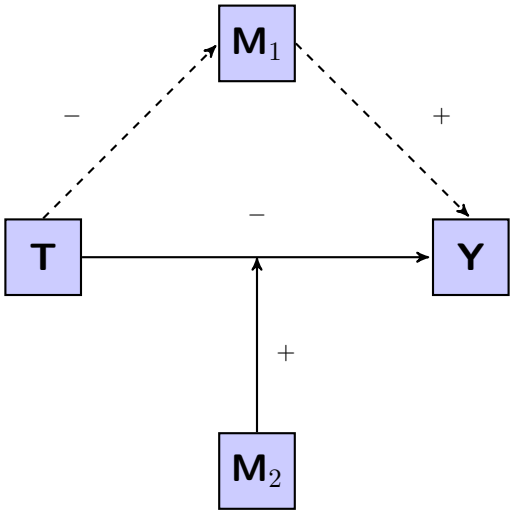

Figure 3.1: The causal mechanism between female gender identity, perceived competence, and perceived student hireability and salary conferral: Female student applicants (T=1) are viewed as less competent (M_1) which then decreases perceived hireability and salary conferral (Y). The effect is more pronounced for individuals with strong implicit gender stereotypes (M_2).

Based on the above considerations and state of research, we designed an experiment to investigate the causal mechanism between female gender, perceived competence, and outcomes relevant to careers in academic science. However, unlike Moss-Racusin et al. (2012) we did

not simply rely on a standard-single experiment design (SED) where subjects are randomly assigned to the treatment and both the level of the mediator and outcome are then measured within both conditions. Instead we applied the parallel encouragement design (PED) recently proposed by Imai et al. (2011). The PED was developed to address the shortcomings of the SED for the investigation of causal mechanisms. Imai and colleagues (2010a) have shown that, for causal mediation effects to be identified, the so-called sequential ignorability assumption (SI) has to hold.

$$Y_i(t', m), M_i(t) \perp\!\!\!\perp T_i | X_i = x, \tag{1}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \tag{2}$$

where $\perp\!\!\!\perp$ denotes statistical independence and $0<\Pr(T_i = t|X_i = x)$ and $0< p(M_i = m|T_i = t, X_i = x)$ for $t = 0, 1$, and all x and m in the support of X_i and M_i, respectively.

In the first part (1) of SI, the treatment assignment (T_i) is assumed to be ignorable (i.e., statistically independent) of potential outcomes (Y_i) and potential mediators (M_i) given the observed pretreatment confounders (x). This assumption usually holds in randomized experiments where treatment is randomly assigned to subjects. In the second part (2) of the assumption, the observed mediator (M_i) is assumed to be statistically independent of potential outcomes (Y_i) given the actual treatment status (t) and pretreatment confounders (x). This implies that no unobserved pre- and posttreatment covariates exist that affect both the mediator and the outcome. Of course, random assignment of the treatment is not sufficient to satisfy the second part of the SI assumption. In fact, researchers can never be certain they have not omitted confounders of the relationship between the mediator and the outcome and thus traditional approaches relying on SEDs and linear structural equation models (LSEM) obscure the SI assumption.

Given these shortcomings, Imai et al. (2011) have proposed the PED which adds a second randomization process to account for the second part of SI. In the PED, subjects are first split into two experiments (see Figure 3.2) which are run in parallel.

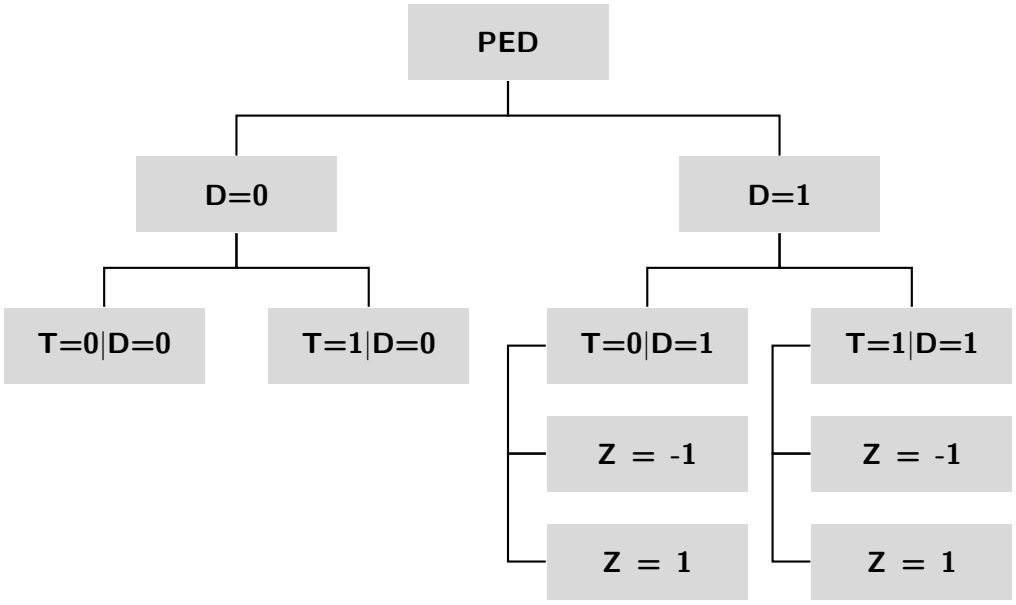

Figure 3.2: Structure of the parallel encouragement design (PED): In the parallel encouragement design, subjects are first randomly assigned to two experiments. In both experiments, subjects are first randomly assigned to the treatment (T=1) and control (T=0) condition. However, in the second experiment (D=1), a random subset of subjects within each condition are encouraged to take a low (Z=-1) or high value (Z=1) of the mediator while subjects in the first experiment (D=0) receive no encouragement.

In the first experiment (D=0), subjects are randomly assigned to the treatment and control condition (= SED). However, in the second experiment (D=1), subjects are first assigned to the treatment and control condition and secondly, within each condition, a random subset of subjects are encouraged to take a low or high value of the mediator. The additional randomization induces an exogenous variation in the mediator (see Figure 3.3) which allows researchers to make inference about the average causal mediation effect (ACME) and average direct effect (ADE) even in the presence of confounders. After all, the randomization of both the treatment and encouragement controls for unobserved confounders represented by the dashed arcs in Figure 3.3.

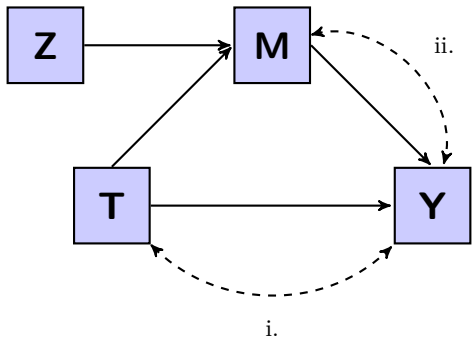

Figure 3.3: Randomized encouragement as an instrument for the mediator: In the parallel encouragement design, the randomization of the treatment (T) allows researchers to control for unobserved confounders of the relationship between the treatment and outcome (Y), which are represented by the dashed arc (i.). The randomized encouragement (Z) induces an exogenous variation in the mediator (M), which allows researchers to make informative inference about the ACME and ADE even in the presence of confounders affecting both the mediator and the outcome (ii.).

Based on these considerations, we decided to conduct a PED for the investigation of the causal mechanism between female gender identity, competence perception, and career-relevant outcomes. Since the mechanism of interest was inherently psychological, we deemed the existence of unobserved confounders to be very likely. Consider, for example, explicit forms of bias such as sexism (e.g., Benson and Vincent 1980; Spence et al. 1973; Swim et al. 1995) or social dominance orientation (Pratto et al. 1994). Both are likely to affect individuals' level of perceived competence of student applicants as well as outcomes that are relevant to women's careers in science. Therefore, we considered the PED to be a better approach allowing for more valid inference than the SED.

3.2 Experimental Design

The PED is suggested to researchers if manipulation of the mediator is imperfect. In our case, manipulation of perceived competence clearly represented such an imperfect manipulation. In fact, it was impossible to instill a low or high value of perceived competence into subjects. For this reason, we encouraged study participants in the second experiment (D=1) to take a low or high score of perceived competence by having them read a slightly modified news story either lauding or criticizing the skill level of engineering students (see the appendix for more details on the encouragements). Since the PED requires a fairly large sample size, we decided to use Amazon's Mechanical Turk (MTurk, http://www.mturk.com) platform for the recruitment and payment of subjects (see the appendix for more information on participants). This procedure allowed us to achieve a large enough sample size and a truly parallel implementation of the experiment. However, it came at the cost of ecological validity since the MTurk population did not represent science faculty. Yet, it allowed us to examine whether

Moss-Racusin et al.'s (2012) results can be generalized beyond science faculty.

As in Moss-Racusin and colleagues' (2012) experiment, we assigned subjects in both experiments to either male or female student application materials for a lab manager position. However, our treatment materials differed from theirs in two important respects. First, we followed recommendations regarding the use of stimulus sampling (e.g., Wells and Windschitl 1999) for studies in which gender represents the manipulation of interest. In such cases, internal validity may be threatened if each of the two gender categories is represented by only one stimulus. This is so because the observed outcomes may in fact be explained by the particular characteristics of the used stimuli instead of gender. Moss-Racusin et al. (2012) applied a between subject design in which faculty participants either read about a male (i.e., "John") or a female student applicant (i.e., "Jennifer") and subsequently evaluated these individuals. Although text materials were identical and names were pretested as equivalent in terms of likeability and recognizeability (Brescoll and Uhlmann 2008), there may still have been unobserved characteristics of the used names that may have accounted for the observed differences in outcomes. For this reason, we chose to create five pairs of identical male and female application materials using four typical male and female names in addition to John and Jennifer (see the appendix for more detailed information on the application materials). Second, in contrast to Moss-Racusin et al. (2012) who focused on academic science in general, our fictitious student applicants all had a background in engineering. We chose this particular discipline based on a large body of literature (e.g., Beggs and Doolittle 1993; Shinar 1975; White et al. 1989; White and White 2006) reporting a strong and prevalent association of engineering with men instead of women. In this context, we deemed it more likely that participants' implicit gender stereotypes would have a relevant impact on their assessment of the respective student applicant. Next, subjects in the second experiment were either assigned to a news story about a survey reporting science faculty's satisfaction or dissatisfaction with engineering students' competence (see the appendix for more information). In a third step, subjects in both experiments indicated their perception of the target student's competence (see the appendix for more information on the measurement), his or her hireability for the laboratory manager position, and the starting salary they would offer him or her (see the appendix for more precise information) before eventually completing the gender-science IAT (see the appendix for further information on the IAT). Note that, unlike in the original study (Moss-Racusin et al. 2012), we did not measure the extent to which participants were willing to mentor the student applicant since these questions were specifically tailored to science faculty. We expected that female gender would exert a negative causal effect on perceptions of competence, hireability and salary conferral (hypothesis A) irrespective of participant gender (hypothesis B), and that these effects would be mediated by perceived competence (hypothesis C). Finally, we expected that subjects' implicit gender bias would moderate results, such that it would be negatively related to evaluations of the female students but unrelated to evaluations of the male applicants (hypothesis D).

3.3 Results

3.3.1 Student gender differences

We first analyzed whether student gender affected perceptions of competence and hireability as well as salary conferral by regressing each variable on student gender, participant gender and their interaction. In contrast to hypothesis A, these analyses revealed (albeit marginally) positive effects of female student gender on salary offer ($\beta = 994.68$, $p = 0.09$) and perceived hireability ($\beta = 0.16$, $p = 0.08$). However, the latter effect was moderated by participant gender (β = -0.29, p<.05). In fact, tests of simple effects showed that male participants rated the female students as slightly more hirable than the identical male students [$t(759)$ = -1.74, $p < 0.05$] while female participants did not have such a preference [$t(716) = 1.25$, $p = 0.11$] (see Table 3.1).

Table 3.1: Means for student competence, hireability, and salary conferral by student gender condition and participant gender

	Male target student				**Female target student**				
	Male participants		**Female participants**		**Male participants**		**Female participants**		
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***d***
Competence	5.18_a	(1.00)	5.22_a	(1.04)	5.26_a	(0.98)	5.16_a	(1.07)	-0.01
Hireability	4.75_a	(1.26)	4.76_a	(1.39)	4.91_b	(1.23)	4.63_{ac}	(1.48)	-0.01
Salary	$34{,}257.43_a$	(8,234.65)	$31{,}811.80_b$	(8,492.98)	$35{,}252.10_c$	(7,785.27)	$33{,}052.49_d$	(8,712.40)	-0.12

Notes: Scales for competence and hireability range from 1 to 7, with higher numbers reflecting a greater extent of each variable. The scale for salary conferral ranges from \$15,000 to \$50,000. Means with different subscripts within each row differ significantly (p<.05). Effect sizes (Cohen's d) represent target student gender differences. Negative effect sizes favor female students. Conventional small, medium, and large effect sizes for d are 0.20, 0.50, and 0.80, respectively. $n_{\text{male student condition}} = 760, n_{\text{female student condition}} = 719$.

Most importantly, both male and female participants favored the female students with respect to salary offer (with female participants generally being more conservative (β = -2,445.63, p<.001)). The mean starting salary offered the female students (\$34,144.65) was significantly [$t(1477)$ = -2.37, $p < 0.01$] higher than the salary offered to the male students (\$33,111.84). Taken together, these results do not support hypotheses A and B. In other words, we could not find evidence of a gender bias working against female student applicants. In contrast to hypotheses A and B, female students were viewed as equally competent as male students by both male and female study participants. In addition, they were viewed as slightly more hirable than male students by male participants and offered more starting salary by both male and female study participants.

3.3.2 Mediation analyses

Instrumental variable estimates

To estimate the direct and indirect effect of student gender, we fitted two-stage least-squares models in which we regressed the outcomes on a binary student gender indicator, coded one for female student applicants and zero for male student applicants, the competence scale, and also controlled for subjects' age, gender, and education (see Table 3.F.2 in the appendix). We instrumented perceived competence with a ternary instrument that was coded minus one if subjects were encouraged to take a low value of perceived competence, zero if they received no encouragement, and one if they were encouraged to take a high value of the mediator. We fitted the first-stage equation by regressing the competence scale on the covariates, the treatment indicator, and the instrument (see Table 3.F.1 in the appendix). We found that the instrument had a strong effect on participants' perceptions of the competence of the respective target student. The F statistic for the weak instrument test is 40, which far exceeds the standard threshold of 10 (Stock and Yogo 2005).

In line with the results from the simple effects analysis, we found that on average the treatment increased salary conferral by 1,033 dollars (with a 95% confidence interval of [217.126, 1849.245]) (see Table 3.2). However, the effect of female student gender was not mediated by participants' perceptions of the competence of the target student. Thus, the female students were not granted more starting salary because they were viewed as more competent than the male student applicants. The same can be said for perceived hireability, where we fail to observe a direct or indirect effect of student gender. Taken together, these results are in striking contrast to those from the original study (Moss-Racusin et al. 2012).

Table 3.2: Estimated causal mediation effects from 2SLS regressions

Outcome	Effect	Estimate
Hireability	Average mediation effect (ACME)	0.001 [-0.108, 0.110]
	Average direct effect (ADE)	-0.001 [-0.082, 0.080]
	Average total effect (ATE)	-0.000 [-0.134, 0.134]
Salary Conferral	Average mediation effect (ACME)	2.258 [-267.429, 271.945]
	Average direct effect (ADE)	1033.185 [217.126, 1849.245]
	Average total effect (ATE)	1035.443 [192.384, 1878.503]

Notes: N = 1479. Entries are estimates from 2SLS regressions. The 95% confidence intervals are based on the delta method. The ATE is the sum of ADE and ACME.

Causal mediation analyses

In a second step, we estimated the average causal mediation (i.e., indirect) effect of student gender using Imai et al.'s (2011) proposed estimation method. To this end, we ran two linear regression models. We first fitted the mediator model where the competence scale was regressed on the treatment (i.e., student gender) and pre-treatment covariates (age, gender, education). We then fitted the outcome models where participants' perceived hireability of the target student and salary conferral were modeled as a function of student gender, competence perception, and the same set of pre-treatment covariates as in the mediator model. We then used the `mediation` software to estimate the average causal mediation effect (ACME) and average direct effect (ADE) of the treatment using the quasi-Bayesian Monte Carlo method based on normal approximation (Imai et al. 2010a) and nonparametric bootstrap (see the appendix for more information on these analyses). Because results did not differ between these two simulation types, we only present results from the quasi-Bayesian Monte Carlo simulations.

Consistent with the previous analyses, we found a positive effect of female student gender on salary conferral (see Table 3.3). Again, however, we found that the effect was not mediated by subjects' perceptions of the competence of the female student applicants. Furthermore, there were no direct or indirect effects of the student gender condition on the perceived hireability of the target student.

Table 3.3: Estimated causal mediation effects of female student gender and perceived competence/likeability

		Mediator	
Outcome	**Effect**	**Competence**	**Likeability**
Hireability	Average mediation effect $\bar{\delta}(0) = \bar{\delta}(1)$	0.007 [-0.106, 0.119]	0.220 [0.126, 0.312]
	Average direct effect $\bar{\zeta}(0) = \bar{\zeta}(1)$	-0.001 [-0.079, 0.085]	-0.211 [-0.316, -0.108]
	Average total effect $\bar{\tau}$	0.007 [-0.128, 0.137]	0.008 [-0.136, 0.147]
Salary Conferral	Average mediation effect $\bar{\delta}(0) = \bar{\delta}(1)$	16.118 [-203.316, 242.118]	486.921 [270.905, 731.703]
	Average direct effect $\bar{\zeta}(0) = \bar{\zeta}(1)$	1063.277 [176.424, 1867.382]	562.157 [-268.349, 1346.031]
	Average total effect $\bar{\tau}$	1079.395 [207.761, 1886.852]	1049.078 [217.603, 1887.426]

Notes: N = 1479. Each cell of the table represents an estimated average causal effect and its 95% confidence interval using the quasi-Bayesian Monte Carlo method with 1000 simulations. The outcomes are participants' perceptions of the hireability of the target students (upper part) and salary conferral (lower part). Results are shown under the no-interaction assumption (i.e., $\bar{\delta}(0) = \bar{\delta}(1)$).

As in Moss-Racusin and colleagues' study (2012), we ran the same analyses with likeability of the student as the mediating variable. Previous research on gender stereotypes (for a review, see Eagly and Mladinic 1994) has shown that people tend to evaluate women more favorably than men. Yet the positive evaluation derives primarily from the ascription of niceness to women and not from characteristics that are considered important for high-status, male-dominated jobs. Consistent with these previous accounts and also the original study (Moss-Racusin et al. 2012), we found greater liking of the female students (M=5.19, SD=1.05) than the male students [(M=4.92, SD=1.11), $t(1477) = -4.86$, $p < 0.001$]. However, in contrast to Moss-Racusin et al. (2012), we found that greater liking of the female students increased the extent to which they were perceived as hirable by study participants as well as the salary they would be offered by them. Concerning the hireability of the student, an otherwise negative direct effect of female student gender (about -0.211 points on the scale of 1 to 7) was offset by an equally sized positive mediation effect (about 0.22 points), which eventually resulted in a total effect of zero (see Figure 3.4). Moreover, the same mechanism accounted for almost half of the positive effect (46% of the total effect) of female student gender on salary conferral. In other words, female students were offered more starting salary by study participants largely because they were viewed as more likeable than their male counterparts.

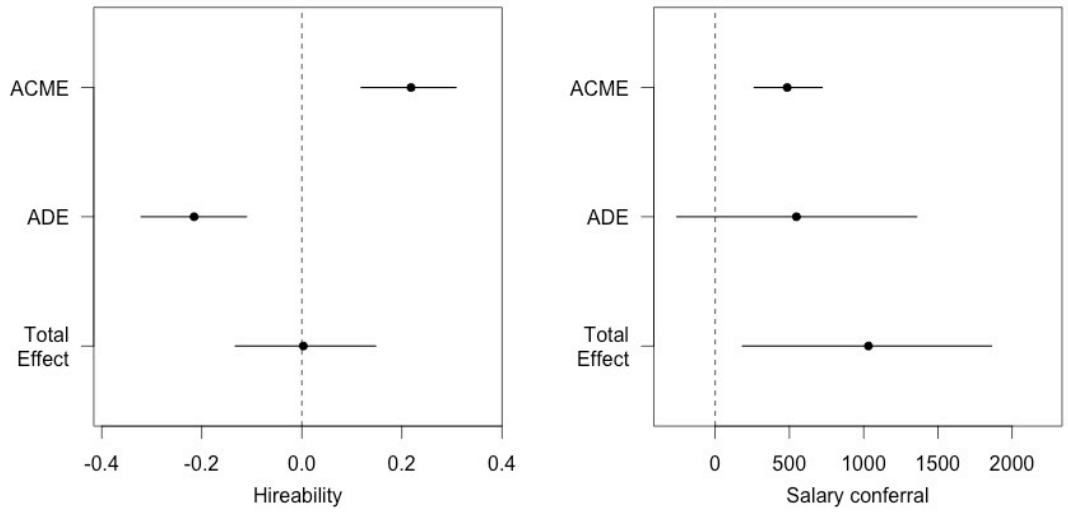

Figure 3.4: Graphical summary of the average mediation effect (ACME), average direct effect (ADE), and the total effect of student gender on perceived hireability (left panel) and salary conferral (right panel).

Finally, we conducted two alternative tests for the robustness of these results to a potential violation of the sequential ignorability (SI) assumption. Note that because the parallel encouragement design accounts for the SI assumption, we only conducted sensitivity analyses for those models in which we included likeability (which was not instrumented) as the mediator. Both types of analysis suggested that the results for salary conferral are more sensitive to a violation of the SI assumption (see the appendix for more information on these analyses). Unfortunately, due to the lack of similar research and applications of such mediation analyses, we are not able to comparatively assess the quality of our results with respect to robustness. However, a comparison with Imai et al.'s (2011) results suggests that at least the findings for hireability are highly satisfactory when it comes to robustness.

Moderated mediation

Preliminary analyses of our gender-science IAT data revealed a slight association between science and male (M=759.44) compared to science and female (M=833.48) among study participants. The mean IAT score (0.20) was significantly different from zero [$t(1462) = 20.45$, $p<.001$] and mean IAT scores did not significantly differ between male (0.19) and female (0.21) subjects [$t(1461) = -0.82$, $p = 0.21$]. Therefore, both male and female participants displayed implicit gender stereotypes, yet only to a relatively weak extent.

Finally, we ran moderated mediation analyses following procedures recommended by Tingley et al. (2014). In particular, we expected that the ACMEs reported above would be

weaker for subjects with strong implicit gender stereotypes. To this end, we compared the ACMEs between participants with IAT scores less or equal to the sample median (0.22) and participants with IAT scores greater than the median. Somewhat surprisingly, these analyses did not provide evidence of moderated mediation (all $p > 0.94$). We therefore reject hypothesis D.

3.4 Discussion

In the present study we sought to replicate previous research (Moss-Racusin et al. 2012) demonstrating a negative effect of gender on female students' chances to pursue academic science careers. Based on the original study by Moss-Racusin et al. (2012), we expected that most of the effect of female gender would be transmitted through participants' perceptions of the competence of students and that the effect would be most pronounced among subjects with high levels of implicit gender stereotypes. Due to the shortcomings of traditional mediation analyses, we applied the recently developed parallel encouragement design (Imai et al. 2011) which takes into account the assumptions that are necessary to identify causal mediation effects. In addition, and in contrast to the original study (Moss-Racusin et al. 2012), we measured implicit gender stereotypes using an implicit association test. The latter represents best practice in contemporary research on implicit cognition and in fact revealed the prevalent association of "male" with "science" and "female" with "liberal arts" as opposed to "female" with "science" and "male" with "liberal arts".

Unlike the original study, we did not per se find a downgrading of female student applicants nor did we find significant mediation effects between student gender, perceived competence and outcomes such as hireability or salary conferral. Instead, we found positive mediation effects between student gender, perceived likeability of the target student and both outcomes. Most notably, participants were willing to offer more starting salary to female students because they viewed them as more likeable than the male student applicants. More precisely, approximately half of the positive direct effect of female gender could be explained by their greater likeability. Interestingly, greater liking of the female students even neutralized the otherwise negative direct effect of female gender identity on participants' perceived hireability of the student. Sensitivity analyses suggested that both results enjoy a satisfactory degree of robustness to unobserved confounding of the relationship between the mediator and the outcome and thus a violation of the sequential ignorability assumption. Another concern regarding our results is that they may be the product of self-presentational concerns among study participants and thus the result of a social desirability bias. However, we do not think that this is the case. First of all, subjects displayed, on average, relatively low levels of implicit gender stereotypes and implicit association tests are generally regarded as non-reactive measures. Thus, it seems unlikely that participants did fake the IAT and that we, as a consequence, underestimated their true levels of implicit gender bias. Rather our results may support the widespread notion of a two-step process of stereotype activation and application. According to a number of research studies (e.g., Bargh 1999; Blair and Banaji 1996; Devine

1989; Kawakami et al. 1998, 2005), implicit stereotypes are automatically activated upon exposure to a member of a particular social group (such as a woman). Hence, an individual is not necessarily aware of the activated stereotypic beliefs and their potential influence on subsequent decision-making. However, when making a judgment, individuals may exert more control over their decision and this may inhibit the application of stereotypes. Secondly, at the end of the study, we asked subjects to report their thoughts on the purpose of the experiment. We excluded all subjects who mentioned gender discrimination, gender bias or similar gender-related aspects and re-ran all analyses finding that results were largely (note that certain results were not significant anymore due to the smaller sample size) unaffected by this data reduction. Thus, we are confident that our results are valid for the American MTurk or internet population.

Our study makes two major contributions. First, we took a step forward in the analysis of causal mechanisms by making use of the PED. To our knowledge, this represents the first application of the PED to a substantive research question. We have demonstrated that researchers can successfully deploy the PED for the study of causal mechanisms provided they have access to a large enough and timely available subject pool. Second, we shed light on the mechanism between gender identity and career-relevant outcomes in the domain of academic science. Other than science faculty members, it seems that the typically young, liberal and Democratic MTurk subjects are less prone to discriminate against female science students and in fact tend to even favor female students because they perceive greater likeability in them than in their male counterparts.

We also acknowledge a major limitation of our study. Due to the constraints of the PED, we were not able to sample individuals who are actually in charge of hiring decisions at science oriented American universities. Our study thus suffers from a certain lack of ecological validity. However, as mentioned above, the average MTurk worker is well-educated (50.2% with a bachelor's degree or higher), fairly liberal in her ideology, and much younger than the sample recruited in the original study. As such, MTurk subjects likely represent the future cohorts of science faculty and our results may thus suggest that gender-based discrimination is likely to become a thing of the past at American universities.

Future research should attempt to apply the PED on a sample of science faculty members to more closely replicate the original study. Also, further work is needed to examine additional mediators of gender effects not only in the domain of academic science but other important spheres of life. Finally, we deem it important to track the development of implicit stereotypes in the long run. Only regular measurements of implicit stereotypes will allow to evaluate the long-term effectiveness of anti-discrimination policies.

References

Amodio, D.M., and Patricia G. Devine. 2006. Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior. *Journal of Personality and Social Psychology*, 91(4), 652-661.

Angrist, J.D., Imbens, G.W., and Donald B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables (with Discussion). *Journal of the American Statistical Association*, 91(434), 444-455.

Bargh, J.A. 1999. The cognitive monster: The case against controllability of automatic stereotype effects. In: Chaiken, S., and Trope, Y. (Eds.), *Dual process theories in social psychology*, 361-382. New York: Guilford Press.

Baron, R.M., and David A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.

Beggs, J.M., and Doolittle. 1993. Perceptions Now and Then of Occupational Sex Typing: A Replication of Shinar's 1975 Study. *Journal of Applied Social Psychology*, 23(17), 1435-1453.

Benson, P.L., and Steven Vincent. 1980. Development and Validation of the Sexist Attitudes Toward Women Scale (SATWS). *Psychology of Women Quarterly*, 5(2), 276-291.

Bertrand, M., Chugh, D., and Sendhil Mullainathan. 2005. Implicit Discrimination. *American Economic Review*, 95(2), 94-98.

Blair, I.V., and Mahzarin R. Banaji. 1996. Automatic and Controlled Processes in Stereotype Priming. *Journal of Personality and Social Psychology*, 70(6), 1142-1163.

Brescoll, V.L., and Eric L. Uhlmann. 2008. Can angry women get ahead? Status conferral, gender, and workplace emotion expression. *Psychological Science*, 19(3), 268-275.

Campbell, B., Schellenberg, E.G., and Charlene Y. Senn. 1997. Evaluating Measures of Contemporary Sexism. *Psychology of Women Quarterly*, 21, 89-102.

Correll, J., Park, B., Judd, C.M., and Bernd Wittenbrink. 2002. The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology*, 83(6), 1314-1329.

Devine, P.G. 1989. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.

Dovidio, J.F., and Samuel L. Gaertner. 2000. Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science*, 11(4), 315-319.

Eagly, A.H. 1987. Sex differences in social behavior: A social role interpretation. Hillsdale, NJ: Erlbaum.

Eagly, A.H., and Antonio Mladinic. 1994. Are People Prejudiced Against Women? Some Answers From Research on Attitudes, Stereotypes, and Judgments of Competence. *European Review of Social Psychology*, 5(1), 1-35.

Foschi, M. 2000. Double standards for competence: Theory and research. *Annual Review of Sociology*, 26(1), 21-42.

Greenwald, A.G., McGhee, D.E., and Jordan L.K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., and Mahzarin R. Banaji. 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.

Heilman, M.E., Wallen, A.S., Fuchs, D., and Melinda M. Tamkins. 2004. Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89(3), 416-427.

Hicks, R., and Dustin Tingley. 2011. Causal mediation analysis. *Stata Journal*, 11(4), 1-15.

Imai, K., Keele, L., and Dustin Tingley. 2010a. A General Approach to Causal Mediation Analysis. *Psychological Methods*, 15(4), 309-334.

Imai, K., Keele, L., and Teppei Yamamoto. 2010b. Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, 25(1), 51-71.

Imai, K., Keele, L., Tingley, D., and Teppei Yamamoto. 2011. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, 105(4), 765-789.

Imai, K., Tingley, D., and Teppei Yamamoto. 2013. Experimental Designs for Identifying Causal Mechanisms. *Journal of the Royal Statistical Society Series A*, 176(1), 5-51.

Kang, J. 2009. Implicit Bias. A Primer for Courts. *Prepared for the National Campaign to Ensure the Ethnic and Racial Fairness of America's State Courts.* Available at: www.americanbar.org/content/dam/aba/migrated/sections/criminaljustice/PublicDocuments/unit_3_kang.authcheckdam.pdf [Accessed: November 21, 2015.]

Kang, J., and Kristin Lane. 2010. Seeing Through Colorblindness: Implicit Bias and the Law. *UCLA Law Rev*, 58(2), 465-520.

Kawakami, K., Dion, K.L., and John F. Dovidio. 1998. Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin*, 24, 407-416.

Kawakami, K., Dovidio, J.F., and Simone v. Kamp. 2005. Kicking the habit: Effects of nonstereotypic association training and correction processes on hiring decisions. *Journal of Experimental Social Psychology*, 41, 68-75.

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., and Jo Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences PNAS*, 109(41), 16474-16479.

Nosek, B.A., Banaji, M.R., and Anthony G. Greenwald. 2002. Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site. *Group Dynamics*, 6(1), 101-115.

Nosek, B.A., Greenwald, A.G., and Mahzarin R. Banaji. 2007. The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In: Bargh, J.A. (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, 265-292. New York: Psychology Press.

Pratto, F., Sidanius, J., Stallworth, L.M., and Bertram F. Malle. 1994. Social Dominance Orientation: A Personality Variable Predicting Social and Political Attitudes. *Journal of Personality and Social Psychology*, 67(4), 741-763.

Reuben, E., Sapienza, P., and Luigi Zingales. 2014. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences PNAS*, 111(12), 4403-4408.

Shinar, E.H. 1975. Sexual Stereotypes of Occupations. *Journal of Vocational Behavior*, 7, 99-111.

Spence, J.T., Helmreich, R., and Joy Stapp. 1973. A short version of the Attitudes Toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, 2(4), 219-220.

Staats, C. 2014. State of the Science: Implicit Bias Review. *Kirwan Institute for the Study of Race and Ethnicity*. Available at: www.kirwaninstitute.osu.edu/wp-content/uploads/2014/03/2014-implicit-bias.pdf [Accessed: November 23, 2015.]

Stock, J.H., and Motohiro Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. In: Stock, J.H., and Donald W.K. Andrews, (Eds.), *Identification and Inference for Econometric Models*, 80-108. New York: Cambridge University Press.

Swim, J.K., Aikin, K.J., Hall, W.S., and Barbara A. Hunter. 1995. Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68(2), 199-214.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Kosuke Imai. 2014. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, 59(5), 1-38.

Wells, G.L., and Paul D. Windschitl. 1999. Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9), 1115-1125.

White, M.J., Kruczek, T.A., Brown, M.T., and Gwendolen B. White. 1989. Occupational Sex Stereotypes among College Students. *Journal of Vocational Behavior*, 34, 289-298.

White, M.J., and Gwendolen B. White. 2006. Implicit and Explicit Occupational Gender Stereotypes. *Sex Roles*, 55, 259-266.

Ziegert, J.C., and Paul J. Hanges. 2005. Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias. *Journal of Applied Psychology*, 90(3), 553-562.

Appendix

3.A Introduction

Apart from the introduction, this appendix includes six sections: After the introduction, I first provide information on the recruitment of participants and descriptive sample statistics. The second section gives an overview of the five treatment stimuli sets that were used in the experiment and also presents an example of the student application materials. In the third section, I present the news articles that served as the encouragement in one of the experiments. The fourth section provides more detailed information on the measurement of all key variables. In the fifth section, I discuss the analytic strategy that we applied and present the results from two-stage least-squares (2SLS) estimation. Finally, in the last section, I present the results from the sensitivity analyses reported in the main article.

3.B Participants and Recruitment

Participants were recruited via Amazon's Mechanical Turk (MTurk, http://www.mturk.com) system, a Web site that enables researchers to publish different kinds of tasks ("Human Intelligence Tasks" (HIT) in Amazon's lingo) and benefit from a large subject pool for online data collection. Only adult people are allowed to register as "workers" and we thus sampled only people eighteen years and older (M=32.32, SD=10.05, range 18-72). We further specified that only workers whose submissions were accepted by task requesters in at least 95 percent of cases were eligible to take our study. While this requirement may slow down the speed of data collection, it helps increase the quality of data by only allowing the most committed workers to take the HIT. Participants were paid $1.00 for completing the study. In total, a sample of 1,479 (761 male, 718 female) adult Americans completed the study in February 2016. The specific ethnic backgrounds were reported as follows: 74% White, 8% Hispanic, 7% African-American, 5% East-Asian, 2% each for South-Asian and multiracial, and 1% each for American Indian/Alaska Native and other. Due to random assignment, participants did not significantly differ between experimental conditions (see Table 3.B.1).

Table 3.B.1: Descriptive statistics for subjects in different experimental conditions

Variable	D00 (n=287)	D01 (n=276)	D10(Z=-1) (n=246)	D10(Z=1) (n=227)	D11(Z=-1) (n=225)	D11(Z=1) (n=218)
Age	32.29 (10.63)	32.17 (9.95)	32.41 (9.52)	33.48 (10.52)	31.75 (10.28)	31.83 (9.20)
Female	0.48 (0.50)	0.51 (0.50)	0.44 (0.50)	0.49 (0.50)	0.48 (0.50)	0.51 (0.50)
Ideology	3.33 (1.73)	3.38 (1.78)	3.42 (1.78)	3.36 (1.78)	3.46 (1.76)	3.25 (1.79)
Partisanship	2.20 (2.03)	2.31 (1.95)	2.38 (2.04)	2.27 (2.00)	2.33 (2.04)	2.21 (2.07)

Notes: Female was a binary indicator (0=male, 1=female); ideology was a 7-point scale (1=very liberal to 7=very conservative); partisanship was a 7-point scale (0=strong democrat to 6=strong republican).

3.C Student Laboratory Manager Application Materials

Since our experiment was a partial replication of Moss-Racusin and colleagues' study (2012), we decided to use similar student application materials. In total, we designed five pairs of male and female treatment stimuli (see Table 3.C.1). Gender was coded 0 for male and 1 for female.

Table 3.C.1: Treatment Stimuli Sets

Pair	Control (= male)	Treatment (= female)	Degree: BA of Science (obtained in)	Age	GPA	GRE score
1	John	Jennifer	Mechanical Engineering (November 2014)	22	3.1	162 verbal 166 quant
2	Peter	Lisa	Technical Engineering (April 2014)	23	3.3	164 verbal 165 quant
3	Paul	Sarah	Systems Engineering (May 2015)	21	3.2	163 verbal 163 quant
4	Greg	Amy	Chemical Engineering (May 2015)	22	3.4	165 verbal 166 quant
5	Jeff	Diana	Industrial Engineering (March 2015)	23	3.2	162 verbal 164 quant

As in the original study by Moss-Racusin et al. (2012) and previous similar research (e.g., Dovidio and Gaertner 2000; Foschi 2000; Heilman et al. 2004), we designed the application materials such that they left enough room for ambiguity and thus subjects' reliance on implicit gender stereotypes when making judgments. In many real-life settings, subjects

are under considerable time pressure when making judgments. In addition, they are mostly provided with rather ambiguous information on the target person to be evaluated. Both time constraints and ambiguous information have been identified as factors supporting the application of implicit (gender) stereotypes (e.g., Bertrand et al. 2005; Eagly 1987; Foschi 2000; Heilman et al. 2004). The text materials were exclusively designed for the purpose of our study and were of similar length. In both experiments, subjects were told they would be assigned to a compilation of application materials of one randomly-selected student. They were instructed to read the description carefully as they would later be asked to assess the qualification of the student to fill a lab manager position. Figure 3.C.1 shows a screenshot of the cover story text. After participants had read the instructions they were assigned to either a male or female stimulus that was randomly drawn from the set of stimuli. Figure 3.C.2 shows the full text of one of the female stimuli.

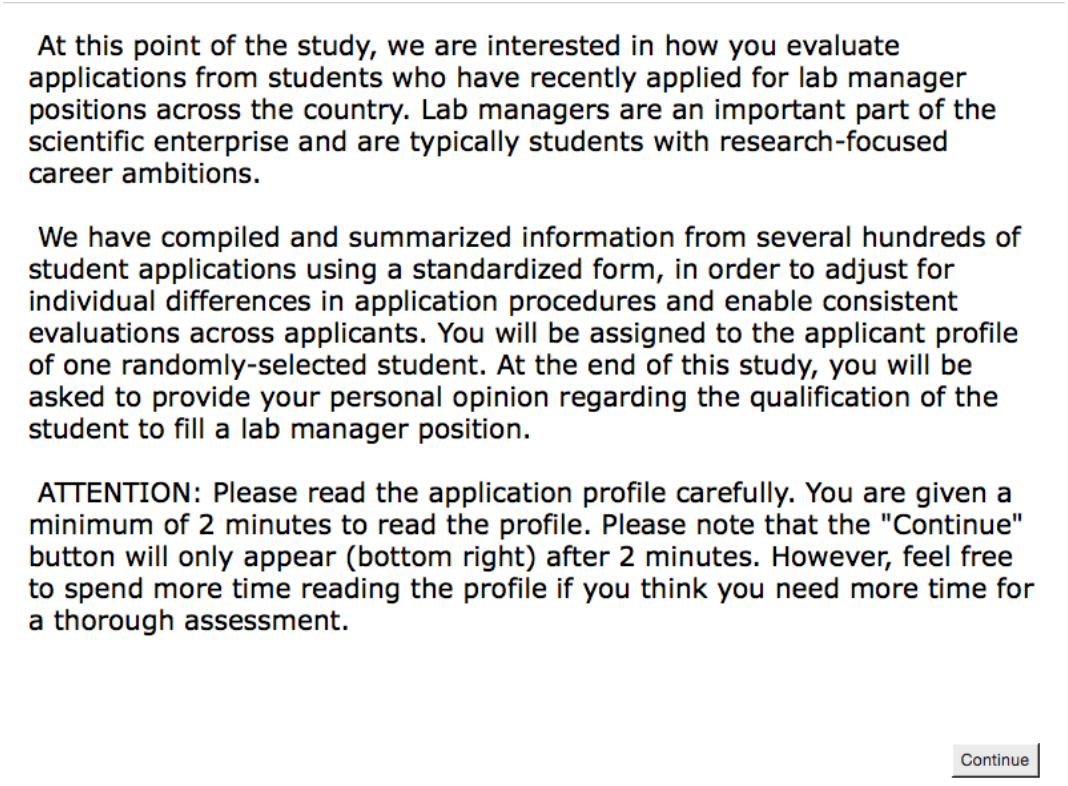

Figure 3.C.1: Cover story text. Participants read this page before they were randomly assigned to one of the student application materials.

DEMOGRAPHICS
Participant ID #: 122
Name: Lisa [redacted]
Gender: Female
Ethnic Background: Caucasian
Age: 23
Degree: Bachelor of Science in Technical Engineering, obtained April 2014 from [redacted] University

BACKGROUND
GPA: 3.3
GRE score: 164 verbal, 165 quant
Awards/honors: President's Service Award, Richard A. Groenendyke Scholarship
Previous research experience: 2 years as a research assistant
Academic standing: appears from Lisa's transcript that she was in good standing upon graduation
Letters of recommendation: 3 (2 from former faculty research supervisors, 1 from intro course professor), all supportive
Future plans: apply to doctoral programs
Extracurricular activities: student government, athletic team
Position sought: Lab Manager
Position duration: 2 years

STATEMENTS/LETTERS
Excerpt from student statement: "I am a very motivated and disciplined person and very committed to an academic research career. After spending a semester at Dr. [redacted]'s chair and another year doing research with Dr. [redacted], I am eager to further improve my knowledge on the development of large-scale IT systems in particular and my research skills in general...as someone planning to apply to doctoral programs, this lab manager position would provide an ideal opportunity to make me competitive for graduate school applications...additionally, I enjoy working as part of an interdisciplinary research team and to benefit from the diverse backgrounds and expertises of all group members. It would be an honor to fill the lab manager position and collaborate with you on future research."

Excerpt from faculty recommendation letter: "Lisa is an excellent candidate for the lab management position...she has great interpersonal skills but might sometimes be too nice...she may not be the best student I have ever taught, but she is surely up there...I am sure her grades would have been better had she not been distracted by her athletic endeavors...with good supervision, she will surely excel."

Figure 3.C.2: Lab manager application materials for one of the female student conditions. The only difference in the corresponding male student condition was that the name "Lisa" was replaced with "Peter", and all female pronouns were replaced with male pronouns.

3.D Encouragements

Because the parallel encouragement design required us to induce a low and high level of perceived competence in the second experiment (D=1), we decided to expose participants to a news article reporting results from a survey on science faculty members' assessment of the skill level of engineering students. In the condition in which subjects were encouraged to take a low value of the mediator, they read the negative version of the article saying that science faculty were very concerned about engineering students' competence for academic research careers. In the other condition, subjects received the opposite message, namely that science faculty were very satisfied with the skill level of engineering students. As can be seen in Figure

3.D.1, the two articles were identical except for the direction of science faculty's expressed opinion. The articles were created based on a recently published news article on the divergent assessments of college students and employers of whether students are well-prepared for the job market (can be accessed here: https://www.insidehighered.com/news/2015/01/20/study-finds-big-gaps-between-student-and-employer-perceptions).

Science faculty [not so] happy with engineering students' competence

Survey finds that engineering students think they are being well-prepared with the skills and qualities needed for scientific careers. Science faculty members [are dubious/agree].

It turns out that college students are being well-prepared for their future careers in engineering science [– at least in their own minds]. Ask science faculty members from American universities, and [it's a very different picture/they agree].

In a recent survey, faculty members and college students were asked a series of similar questions about career preparation. They [could/can] be [scary/promising] reading for many students eager to pursue a career in engineering science. Consistent with past surveys, this one found that science faculty are [very concerned/very satisfied] [about/with] new graduates [having a range of/and their] skills in areas like communication and team work. In fact, students are [more than twice] as likely as employers to rank themselves as prepared in areas such as oral and written communication, critical thinking, and being innovative/creative. "When it comes to the types of skills and knowledge that science faculty feel are most important to success in academic science, large majorities of researchers [do not] feel that recent college graduates are well prepared. This is particularly the case for applying knowledge and skills in real-world settings, critical thinking skills, and written and oral communication skills. But even in the areas of ethical decision-making and working with others in teams, many faculty participants [do not] give graduates high marks," the survey report says.

Figure 3.D.1: News articles in the [low/high] encouragement conditions of the second experiment (D=1)

3.E Measures

Perceived competence

We asked subjects to assess the target student's competence using the same items that were used in the original study of Moss-Racusin et al. (2012). Subjects answered these questions on a 1 (not at all) to 7 (very much) scale. The three items were: (i) Did the applicant strike you as competent? (ii) How likely is it that the applicant has the necessary skills for this job? (iii) How qualified do you think the applicant is? ($\alpha = 0.89$). The three items were averaged to form the competence scale, with higher values indicating greater levels of perceived competence.

Implicit gender stereotypes

Unlike Moss-Racusin et al. (2012), we decided to use a measure that truly captures implicit (i.e., automatic) gender stereotypes. Implicit association tests (IAT) (Greenwald et al. 1998) have become best practice when it comes to measuring automatic beliefs. Of particular importance to our study was evidence from a gender-science IAT that demonstrated the prevalent implicit association of "male" with "science" and "female" with "liberal arts" (Nosek et al. 2002). We expected that these implicit stereotypes would moderate the effect of female gender identity on career-relevant factors such as hireability and salary offer. In both experiments, participants first completed (see Table 3.E.1) a practice block in which male stimuli (e.g., man) were either assigned to the left-key response (i.e., "E") and female stimuli (e.g., woman) to the right-key response (i.e., "I") or vice versa. Next, subjects completed 20 trials in which stimuli pertaining to "science" (e.g., engineering) were assigned to the left-key response and stimuli pertaining to "liberal arts" (e.g., history) to the right-key response. After these practice blocks subjects completed two blocks of 20 and 40 trials in which "male" and "female" were - depending on their position in the first block - either paired with "science" or "liberal arts". In the fifth block, the position of the concepts "male" and "female" was reversed to allow subjects to get accustomed to the new position. Finally, subjects completed two more blocks in which the pairings of "male" and "female" with "science" and "liberal arts" were reversed in comparison to blocks three and four. Block order was counterbalanced across participants and the stimuli within each block were presented in random order. All four categories were represented by eight stimuli (see Table 3.E.2) that were already used in previous applications of the gender-science IAT (Nosek et al. 2002). Figure 3.E.1 provides a sample screenshot of the IAT.

Table 3.E.1: Sequence of trial blocks in Gender-Science IAT

Block	No. of Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Concept discrimination	Male	Female
2	20	Attribute discrimination	Science	Liberal Arts
3	20	Practice combined task	Science / Male	Liberal Arts / Female
4	40	Combined task	Science / Male	Liberal Arts / Female
5	20	Reversed concept discrimination	Female	Male
6	20	Practice reversed combined task	Science / Female	Liberal Arts / Male
7	40	Reversed combined task	Science / Female	Liberal Arts / Male

Note: The IAT score is based on data from blocks 3, 4, 6, and 7. The order of the combined blocks was counterbalanced across participants. Stimuli were presented in random order.

Table 3.E.2: Stimuli used in Gender-Science IAT

Category	Stimuli
Male	Male, Man, Boy, Brother, He, Him, His, Son
Female	Female, Woman, Girl, Sister, She, Her, Hers, Daughter
Science	Chemistry, Physics, Biology, Biophysics, Engineering, Astronomy, Biochemistry Neuroscience
Liberal Arts	Philosophy, Arts, Humanities, History, Spanish, English, Latin, Music

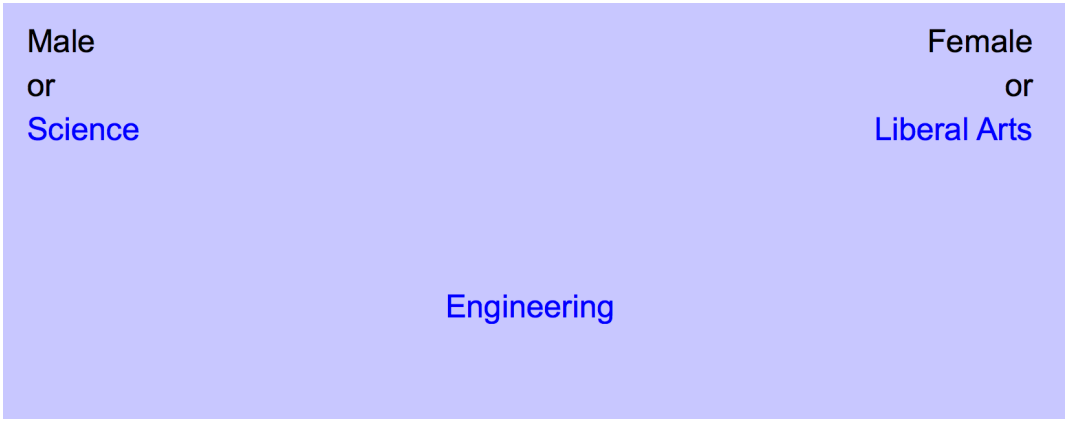

Figure 3.E.1: Screenshot of the IAT

The IAT scores were computed in line with the improved scoring algorithm D2 (Greenwald et al. 2003). Therefore, all trials with response latencies less than 400 ms or greater than 10 seconds were discarded from the analyses and a built-in error penalty was used in the case of incorrect responses. The scores can take on values between -2.0 and 2.0 and thus inform about both the strength and direction of one's implicit association. Scores were interpreted as follows:

IAT $D \leq$ -0.65 = "a strong association of male with liberal arts and female with science compared to male with science and female with liberal arts"

-0.65 $<$ IAT $D \leq$ -0.35 = "a moderate association of male with liberal arts and female with science compared to male with science and female with liberal arts"

-0.35 $<$ IAT $D \leq$ -0.15 = "a slight association of male with liberal arts and female with science compared to male with science and female with liberal arts"

-0.15 < IAT D < 0.00 = "little to no association of male with liberal arts and female with science compared to male with science and female with liberal arts"

0.00 < IAT $D \leq$ 0.15 = "little to no association of male with science and female with liberal arts compared to male with liberal arts and female with science"

0.15 < IAT $D \leq$ 0.35 = "a slight association of male with science and female with liberal arts compared to male with liberal arts and female with science"

0.35 < IAT D < 0.65 = "a moderate association of male with science and female with liberal arts compared to male with liberal arts and female with science"

IAT $D \geq$ 0.65 = "a strong association of male with science and female with liberal arts compared to male with liberal arts and female with science"

Dependent variables

Hireability
The extent to which subjects viewed the respective student as hireable was measured with the same items as in Moss-Racusin and colleagues' (2012) study. Subjects answered the following three items on a 1 (not at all likely) to 7 (very likely) scale: (i) How likely would you be to invite the applicant to interview for the laboratory manager job? (ii) How likely would you be to hire the applicant for the laboratory manager job? (iii) How likely do you think it is that the applicant was actually hired for the laboratory manager job he/she applied for? (α = 0.90). Items were averaged to form the student hireability scale, with higher values indicating greater perceived hireability.

Salary conferral
As in the original study (Moss-Racusin et al. 2012), we measured salary conferral with one item: If you had to choose one of the following starting salaries for the applicant, what would it be? Participants could use the following scale to indicate the salary they deemed appropriate for the respective student applicant: 1 ($15,000), 2 ($20,000), 3 ($25,000), 4 ($30,000), 5 ($35,000), 6 ($40,000) 7 ($45,000) 8 ($50,000). Collapsed across all conditions, the average recommended salary was $33,613.93 (SD=8,400.84), with a range of $15,000 to $50,000.

Likeability
Finally, we also measured the extent to which study participants liked the student applicant using three items on a 1 (not at all) to 7 (very much) scale as in the original study (Moss-Racusin et al. 2012). The items were: (i) How much did you like the applicant? (ii) Would you characterize the applicant as someone you want to get to know better? (iii) What do you think: Would the applicant fit in well with other laboratory members? The items were

averaged to create the likeability scale, with higher numbers reflecting greater liking of the respective target student ($\alpha = 0.81$).

3.F Analytic Strategy

We applied a twofold strategy to analyze the data from the parallel encouragement design (PED). First, we analyzed the data using the instrumental variable approach (Angrist et al. 1996). In this setting, the randomized encouragement represents the instrument for the mediator. Data is typically analyzed using two-stage least-squares (2SLS) estimation. The two stages are written as follows (Imai et al. 2011):

$$M_i = \alpha_1 + \beta_1 T_i + \lambda Z_i + \xi_1^T X_i + \epsilon_{i1}$$

$$Y_i = \alpha_2 + \beta_2 T_i + \gamma M_i + \xi_2^T X_i + \epsilon_{i2}$$

It follows that the product of coefficients $\beta_1\gamma$ yields the estimate of the average causal mediation effect (ACME) while β_2 is the estimate of the average direct effect (ADE). We fit the first stage equation (see Table 3.F.1) by regressing perceived competence (M) on the covariates (X), the binary treatment (T) indicator (0 = male, 1 = female), and encouragement (Z) status (-1 if mediator is encouraged down, 0 if no encouragement, and 1 if encouraged up). In the second stage (see Table 3.F.2), the outcomes are regressed on the predicted values of the mediator, the treatment, and pre-treatment covariates (X).

Table 3.F.1: First Stage Regression Estimates

Outcome:	**Competence**
Encouragement: Down	-.38***
	(.06)
Encouragement: Up	.19**
	(.06)
Female Student Gender (0/1)	.00
	(.05)
Age	-.00
	(.00)
Female (0/1)	-.04
	(.05)
Education: High School	-.77*
	(.32)
Education: Some College	-.84**
	(.31)
Education: Bachelor's degree or higher	-.95**
	(.30)
F-test	40.06

Notes: N=1479. Ordinary least squares regression of perceived competence on student gender (the treatment) and ternary instrument (-1 if mediator is encouraged down, 0 if no encouragement, and 1 if encouraged up). Robust standard errors in parentheses. *** p<.001, ** p<.01, * p<.05.

Table 3.F.2: 2SLS Estimates of the Effect of Female Student Gender on Perceived Hireability and Salary Conferral

	(1)	(2)
Outcome:	**Hireability**	**Salary Conferral**
Female Student Gender (0/1)	-.00	1033.19*
	(.07)	(431.30)
Competence	1.07***	2655.87**
	(.15)	(963.96)
Age	-.00	-60.36**
	(.00)	(20.98)
Female (0/1)	-.11	-2204.44***
	(.29)	(433.27)
Education: High School	-.11	-440.97
	(.29)	(4268.91)
Education: Some College	-.07	129.68
	(.27)	(4212.85)
Education: Bachelor's degree or higher	-.08	330.11
	(.28)	(4228.69)

Notes: Instrumental variables regression of outcomes (1) and (2) on female student gender and perceived competence (instrumented by the encouragement). Robust standard errors in parentheses. The following covariates are used as controls: age, gender, education. *** p<.001, ** p<.01, * p<.05.

The validity of the causal inference from 2SLS estimation hinges on two assumptions that have to hold simultaneously. First, the instrument (i.e., encouragement) is assumed to affect the outcome only through the mediator (i.e., perceived competence). While known as the exclusion restriction in the instrumental variable method, the assumption is referred to as consistency in the PED (Imai et al. 2013). Formally, for all t = 0, 1 and z, z' = -1, 0, 1 (where -1 = negative report on students' competence, 0 = no reading, 1 = positive report on students' competence) ...

$$Y_i(t, M_i(t, z), z) = Y_i(t, M_i(t, z'), z') \text{ if } M_i(t, z) = M_i(t, z')$$

Second, it is assumed that the encouragement or instrument (Z) monotonically affects the mediator. In other words, there are no "defiers" who behave exactly opposite to the encouragement. Formally, for t = 0, 1 and z, z' = -1, 0, 1 ...

$$M_i(t,1) \geq M_i(t,0) \geq M_i(t,-1)$$

How plausible were these assumptions in the context of our study? As in most applications of instrumental variables, monotonicity was the less problematic assumption. In fact, it was rather unlikely that subjects' level of perceived competence would increase upon exposure to a news article reporting that science faculty were very unhappy with engineering students' competence. Similarly, it was unlikely that study participants would perceive the student applicants as less competent as a function of reading the positive version of the article. However, as is typically the case with instrumental variables, consistency is a relatively strong assumption. This is particularly true when researchers are interested in a psychological mechanism (Imai et al. 2011). For example, having subjects read a news article about the results of a survey on students' skill levels may have a direct effect on the outcomes if the task changes other assessments or perceptions which in turn affect perceived hireability and salary conferral.

As a second analysis, we ran causal mediation analyses using the `mediation` package for STATA (Hicks and Tingley 2011) and R (Tingley et al. 2014). As outlined in the main article, we conducted both causal mediation analysis based on quasi-Bayesian Monte Carlo simulation (Imai et al. 2010a) and nonparametric bootstrap for variance estimation. In both analyses we used a number of 1000 simulations and robust standard errors. Furthermore, for each analysis, we tested for an interaction between the treatment and mediator variable using the test recommended by Tingley et al. (2014). For each model, we found that the ACMEs were not significantly different as a function of treatment status and all our analyses were thus based on the no interaction assumption. Note that results from both quasi-Bayesian Monte Carlo simulations and nonparametric bootstrap analyses were largely identical.

3.G Sensitivity Analyses

Following recommendations by Imai and colleagues (2011) we ran two types of sensitivity analysis. First, as shown in Imai et al. (2010a), SI implies zero correlation between the error terms of the mediator (1) and outcome model (2). Therefore, one way of testing the robustness of the ACME is to vary the value of this correlation (ρ) and identify the point at which the ACME changes sign (Imai et al. 2010a,b, 2011). Using the function `medsens` from the `mediation` package yielded the results presented in the top row of Figure 3.G.1. We find that for perceived hireability, the estimated ACME equals zero when ρ equals 0.65. If we take into account sampling uncertainty, we find that the 95% confidence intervals for the ACME include zero when ρ exceeds 0.63. Thus, an unobserved confounder would have to affect both likeability and hireability in the same direction and make the correlation between the two error terms greater than 0.63 for us to reject a positive ACME. Likewise, we find that for salary conferral, the estimated ACME equals zero when ρ equals 0.24. These results imply that the

ACME estimate for hireability is much more robust to unobserved confounding. A second sensitivity analysis examines the portion of the residual and total variance in the mediator and outcome model that needs to be explained by an unobserved confounder for the ACME to become zero. If it turns out that only a small portion of the variance in the mediator and outcome must be explained by some omitted variable then results must be considered weak. The results of these analyses are presented in the bottom row of Figure 3.G.1. For perceived hireability, we find that the true ACME changes sign if the product of these proportions is greater than 0.23 and the confounder affects both likeability and hireability in the same direction. For example, if an unobserved confounder explains more than 58% of the variance in likeability and more than 40% of the variance of the hireability scale, then the true ACME is negative. In other words, an unobserved confounder must explain as much as 48% ($\simeq \sqrt{0.23}$) of the variance in the mediator and outcome model for the true ACME to become negative. Turning our attention to the results for salary conferral we find that an unobserved confounder must at least explain 22% of the total variance in both variables for the true ACME to be negative. Therefore, we again conclude that results for salary conferral are less robust to a potential violation of the SI assumption.

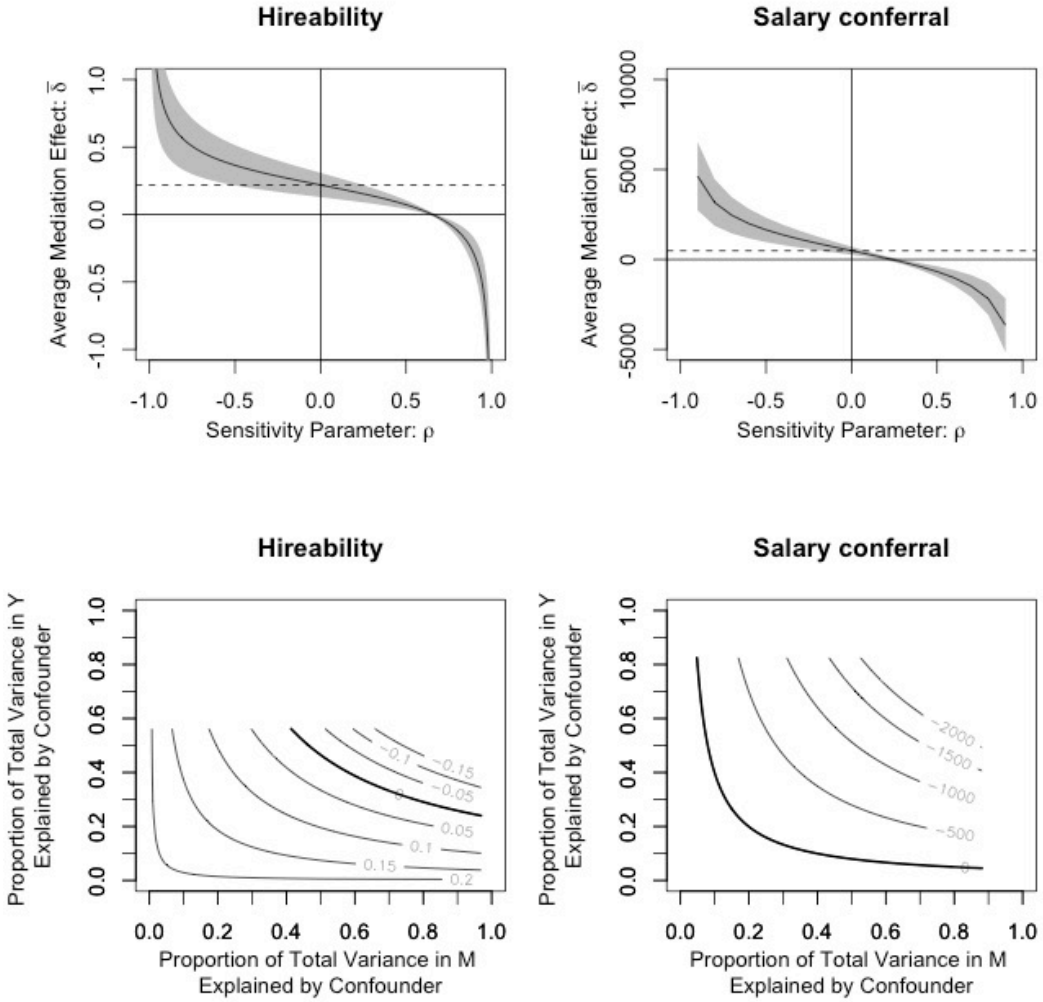

Figure 3.G.1: Sensitivity analyses for the causal mechanism between student gender, likeability and perceived hireability (left) and salary conferral (right). The top row plots the ACME as a function of the sensitivity parameter ρ, which is the correlation between the error terms in the mediator and outcome regression models. The dashed line represents the estimated ACME when the sequential ignorability assumption is made. The shaded areas represent the 95% confidence interval for the mediation effects at each value of ρ. The bottom row of plots provide the alternative formulation based on the decomposition of variances. Here, the contours represent the true ACME plotted as a function of the proportion of the total mediator variance (horizontal axis) and the total outcome variance (vertical axis), that are each explained by the unobserved confounder included in the corresponding regression models. Here the unobserved confounder is assumed to affect the mediator and outcome in the same direction. The sensitivity analyses suggest that the results for salary conferral are more sensitive to an omitted variable that affects the perceived likeability of the target student and salary offer in the same direction.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX typesetting system. The template used for this document has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.